The effects of p53 gene inactivation on mutant proteome expression in a human melanoma cell model

Elsevier Editorial System(tm) for BBA - General Subjects

Manuscript Draft

Manuscript Number: BBAGEN-20-251R1

Title: The effects of p53 gene inactivation on mutant proteome expression in a human melanoma cell model

Article Type: Regular Paper

Keywords: cancer; p53; protein mass spectrometry; proteogenomics; proteomics

Corresponding Author: Professor Ted R Hupp, PhD

Corresponding Author's Institution: University of Edinburgh

First Author: Jakub Faktor

Order of Authors: Jakub Faktor; Giuseppa Grasso; Filip Zavadil; Małgorzata Kurkowiak; Marcos Y Mayordomo; Sachin Kote; Ashita Singh; Li Ruidong; Robert O'Neill; Petr Muller; David Goodlett; Borek Vojtesek; Ted R Hupp, PhD

Abstract: ABSTRACT
Background: The identification of mutated proteins in human cancer cells-termed proteogenomics requires several technologically independent research methodologies including genomics, RNA sequencing, and mass spectrometry. Any one of these methodologies are not optimized for identifying potential mutated proteins and any one output fails to cover completely a specific landscape.
Methods: An isogenic melanoma cell with a p53-null genotype was created by CRISPR/CAS9 system to determine how p53 gene inactivation affects mutant proteome expression. A mutant peptide reference database was developed by comparing two distinct DNA and RNA variant detection platforms using these isogenic cells. Chemically fractionated tryptic peptides from lysates were processed using a TripleTOF 5600+ mass spectrometer and their spectra were identified against this mutant reference database.
Results: Approximately 190 mutated peptides were enriched in wt-p53 cells, 187 mutant peptides were enriched in p53-null cells, with an overlap of 147 mutated peptides. STRING analysis highlighted that the wt-p53 cell line was enriched for mutant protein pathways such as CDC5L and POLR1B, whilst the p53-null cell line was enriched for mutated proteins comprising EGF/YES, Ubiquitination, and RPL26/5 nodes.
Conclusion: Our study produces a well annotated p53-dependent and p53-independent mutant proteome of a common melanoma cell line model. Coupled to the application of an integrated DNA and RNA variant detection platform (CLCbio) and software for identification of proteins (ProteinPilot), this pipeline can be used to detect high confident mutant proteins in cells.
General significance: This pipeline forms a blueprint for identifying mutated proteins in diseased cell systems.

Response to Reviewers: Reviewer #1

Major comments:
This is an interesting study, which provides a pipeline to study the difference in gene expression levels and mutant proteins between p53-positive and p53-negative melanoma cells. These isogenic cell lines were generated using CRISPR/Cas9 genome editing of the parental A375 cells. However, it is not clear from the description whether the CRiSPR/Cas9-introduced mutation in the p53 gene before ARG175 produced clean knockout. Depending on the p53 antibody used in the study it may not recognize the amino-terminus of p53. Thus, this relevant information about p53 antibodies used in the present work should be included in the text.

RESPONSE: On the page 26 we added a new figure 7F to show that in both HCT116 and A375 p53-null cell lines, there is no p53 protein detected using a monoclonal antibody that binds to the N-terminal domain of p53, DO-1. As the reviewer correctly suggested, it is possible that the RNA edits around codon 175 produced an out of frame stop codon that resulted in truncated p53 protein. The use of the monoclonal antibody DO-1 confirms that the gene edit at codon 175 position does not produce a truncated p53 protein and that the A375-p53 edited cell line is a true "p53 protein null cell". We discuss these findings on the page 8. We added a section describing how we validated loss of p53 in materials and methods (page 11).

Minor comments:
there are several typos in the text, e.g. how p53 gene deletion effects mutant proteome expression (needs to be changed to affect).

RESPONSE: We thoroughly searched entire text for typos and corrected them.

BBAGEN-20-251 Revision Cover Letter

July 29, 2020

Lee Graves, Executive Editor
BBA - General Subjects

Dear Lee,
we are pleased to submit a revised version of our manuscript entitled "The effects of p53 gene inactivation on mutant proteome expression in a human melanoma cell model" (BBAGEN-20-251). Thank you for giving us the opportunity to revise and resubmit this manuscript. We are submitting this revision upon deadline August 31, 2020.

We thank you and reviewers for careful review and comments that will lead to improvement of the original version of the manuscript. We have incorporated the suggested changes into manuscript and we respond to all issues raised by reviewers below. We are looking to cooperate with you and reviewers to move this submission closer to publication in BBA - General Subjects.

Reviewer #1
Major comments:
This is an interesting study, which provides a pipeline to study the difference in gene expression levels and mutant proteins between p53-positive and p53-negative melanoma cells. These isogenic cell lines were generated using CRISPR/Cas9 genome editing of the parental A375 cells. However, it is not clear from the description whether the CRiSPR/Cas9-introduced mutation in the p53 gene before ARG175 produced clean knockout. Depending on the p53 antibody used in the study it may not recognize the amino-terminus of p53. Thus, this relevant information about p53 antibodies used in the present work should be included in the text.

RESPONSE: On the page 26 we added a new figure 7F to show that in both HCT116 and A375 p53-null cell lines, there is no p53 protein detected using a monoclonal antibody that binds to the N-terminal domain of p53, DO–1. As the reviewer correctly suggested, it is possible that the RNA edits around codon 175 produced an out of frame stop codon that resulted in truncated p53 protein. The use of the monoclonal antibody DO–1 confirms that the gene edit at codon 175 position does not produce a truncated p53 protein and that the A375-p53 edited cell line is a true "p53 protein null cell". We discuss these findings on the page 8. We added a section describing how we validated loss of p53 in materials and methods (page 11).

Minor comments:
there are several typos in the text, e.g. how p53 gene deletion effects mutant proteome expression (needs to be changed to affect).

RESPONSE: We thoroughly searched entire text for typos and corrected them.

Sincerely, Ted Hupp

Highlights

1. The study of mutated proteomes in cancer cells is at its infancy
2. A proteogenomics platform is developed that exploits DNA and RNA variant identification
3. Mass spectrometry was used to identify and validate mutant protein expression
4. P53-dependent mutant proteome networks were identified
5. Proteogenomics pipelines form a blueprint for identifying mutated proteins in diseased systems

The effects of p53 gene inactivation on mutant proteome expression in a human melanoma cell model

Jakub Faktor[2], Giuseppa Grasso[1+], Filip Zavadil[2], Małgorzata Kurkowiak[3], Marcos Yébenes Mayordomo[1,3], Sachin Kote[3], Ashita Singh[1], Li Ruidong[1,5], J. Robert O'Neill[1, 6], Petr Muller[2], David Goodlett[3,4], Borek Vojtesek[2], and Ted Hupp[1,2,3]

[1]University of Edinburgh, Institute of Genetics and Molecular Medicine, Edinburgh, UK; [2]Research Centre for Applied Molecular Oncology (RECAMO), Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic; [3]University of Gdansk, International Centre for Cancer Vaccine Science, ul. Wita Stwosza 63, 80-308 Gdansk, Poland; [4]School of Pharmacy, University of Maryland, Baltimore, USA ICCVS; [5]Department of Gastrointestinal Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China. [6]University of Cambridge, Cambridge, UK. + current address, Institut de Génétique Humaine, CNRS, University of Montpellier, Gene Regulation Laboratory, 141 rue de la cardonille, Montpellier, France.

Running title: An isogenic p53-null melanoma cell model for use in mutant proteomics

Keywords: cancer, p53, protein mass spectrometry, proteogenomics, proteomics

Correspondence, vojtesek@mou.cz , ted.hupp@ed.ac.uk

ABSTRACT

Background: The identification of mutated proteins in human cancer cells-termed proteogenomics, requires several technologically independent research methodologies including DNA variant identification, RNA sequencing, and mass spectrometry. Any one of these methodologies are not optimized for identifying potential mutated proteins and any one output fails to cover completely a specific landscape.

Methods: An isogenic melanoma cell with a p53-null genotype was created by CRISPR/CAS9 system to determine how p53 gene inactivation affects mutant proteome expression. A mutant peptide reference database was developed by comparing two distinct DNA and RNA variant detection platforms using these isogenic cells. Chemically fractionated tryptic peptides from lysates were processed using a TripleTOF 5600+ mass spectrometer and their spectra were identified against this mutant reference database.

Results: Approximately 190 mutated peptides were enriched in wt-p53 cells, 187 mutant peptides were enriched in p53-null cells, with an overlap of 147 mutated peptides. STRING analysis highlighted that the wt-p53 cell line was enriched for mutant protein pathways such as CDC5L and POLR1B, whilst the p53-null cell line was enriched for mutated proteins comprising EGF/YES, Ubiquitination, and RPL26/5 nodes.

Conclusion: Our study produces a well annotated p53-dependent and p53-independent mutant proteome of a common melanoma cell line model. Coupled to the application of an integrated DNA and RNA variant detection platform (CLCbio) and software for identification of proteins (ProteinPilot), this pipeline can be used to detect high confident mutant proteins in cells.

General significance: This pipeline forms a blueprint for identifying mutated proteins in diseased cell systems.

INTRODUCTION

Next-generation genome sequencing technologies have revolutionized our understanding of the molecular nature of cancer [1] [2]. Parallel innovations in combinatorial chemistry, crystallography, high-throughput drug screening, transgenesis, and computational science have rapidly generated hundreds of promising targeted drug leads. However, despite this increased R&D, the number of effective drugs reaching the clinic is in steady decline [3]. A technical problem is the lack of robust age-dependent sporadic immune competent models of human cancer that predicts human toxicity and response [4]. Another major hurdle is that sporadic cancers are multi-gene diseases thus minimizing the likelihood of finding a common (set of) drug(s) to improve patient welfare.

Whole genome cancer sequencing has defined the strikingly patient-specific cancer bar code thus highlighting the unique genetic signature of any given tumour [5] [6]. However, the vast majority of anti-cancer medicines target wild-type proteins, although there are ever emerging successes in targeting mutated kinases with effective drug leads [7]. This presents an opportunity to develop precision, personalized therapeutics based on expressed, mutant proteins. Such mutant proteins could inform target pathway choice for the development of novel Biologics that target the mutated cancer landscape in a patient-specific manner [8]. Understanding the expression of mutant proteins could also form a platform for the development of mutant neoantigen based anti-cancer vaccines, that could be based on synthetic proteins [9], dendritic cells primed with neoantigens [10], nucleic acids such as RNA [11], or synthetic viral vectors [12]. However, the study of mutant proteomes is only in its infancy. The relative difficulty in this task is that building mutant proteomes requires integration of robust and user-friendly methods linking the fields of informatics, mass spectrometry, and cancer biology. This task is not necessarily trivial.

Computational methodologies as applied to the cancer research field are emerging as approaches to define the expressed, mutated genome. There are several challenges with optimizing platforms that integrate DNA sequencing, RNA sequencing, and mass spectrometric datasets [13]. One overall challenge is integrating these molecular data into one pipeline; and depends on the variant detection platform used for DNA sequencing; the algorithms for defining mutations with RNA sequencing datasets in exons, non-coding RNAs, and introns [14]. For example, the expression of intron encoded mutant peptides is almost completely unexplored at a systems biology level [15], as are cancer-specific RNA edits and tumour-specific spliced mRNAs that create cancer-associated polypeptide epitopes. In addition, different mass spectrometers, sample preparation and pre-fractionation methods, coupled to tumor heterogeneity, result in an incomplete understanding in the source and extent of expressed mutated proteins using cancer-specific DNA and/or RNA sequencing reference databases.

There are several types of platforms recently developed for integrating DNA, RNA and protein data integration. ‘Proteoformer’ produces a complete protein synthesis database that can be used to identify peptides with mass spectrometry through the use of ribosome profiling [16]. The limitation of this approach is that living cells are required to isolate bioactive ribosomes and this method might not be conducive to frozen tissues from the clinic. Methods have been established for automating spliced variants in cell systems which is especially powerful in cancer genomes where there might be DNA fusions, cancer-specific splicing, and trans-splicing [17] [18]. Spliced variant detection algorithms are always improving, especially those that capture the pathological, heterogeneous splicing specific to cancer cells [19]. Modification of R-packages iterates innovations in identifying expressed mutated genes [20]. Translation toolkits have been generated that aim to produce a theoretical total polypeptide space of a genome using RNAseq that captures

six-frame genome translation [21]. These examples highlight the types of several bespoke algorithms that generate cutting-edge information on the proteogenomic landscape. As the diversity in software and computational tools tend not to be benchmarked against each other, end-user compatibility, especially crossing different disciplines, can be limited or non-accessible.

In this report, we focus first on benchmarking two distinct DNA and RNA variant calling platforms towards identifying the mutant proteome landscape in a biological human cancer cell model. One of the variant detection platforms; *CLCbio*, is an integrated DNA and RNA variant identification software platform. This software has been used previously in variant detection using human cancer tissue [22] [23] although it was not benchmarked against more classic variant calling platforms. The utility of the *CLCbio* application is that it is a tool not requiring computational coding and can therefore open the door for life-scientists to identify tissue or cell-specific genomic variants relevant to the biological system of interest. Such an application for life-scientists that does not require coding already exists for research in the mass spectrometric field, such as *Proteome Discoverer* [24]. The most common coding-dependent genome analysis toolkits for DNA variant detection are platforms such as Varscan and Mutect [25]. In this current study, we not only benchmark both the *CLCbio* software and *Varscan2* platforms, we also use these data to create a mutant reference dataset to define mutated proteins in a key cancer cell model. We focus our biological question on asking an emerging topic in the cancer research field; how does inactivation of a major cancer-associated gene impact on the mutant proteome landscape? We answer this question by creating an isogenic melanoma cell panel with a p53 gene inactivation to define changes in the mutant proteome as function of p53 gene inactivation.

RESULTS

Using CLCbio and Varscan2 DNA variant detection platforms to develop a mutant genomic reference database using a human melanoma cancer cell line model.

The human melanoma A375 cell line has classically been used as a model to study regulation of wt-p53 activity in response to DNA damage [26] [27]. Next generation DNA sequencing of the A375 wt-p53 containing melanoma genome was performed in order to annotate its genome. This also produces a community resource that can be used to develop proteogenomics tools and pipelines for use in studying both mutated proteome expression and/or neoantigen production. The majority of next-generation data analysis using DNA variant detectors derived from *Varscan* or *Mutect* requires computational coding skills [25]. By contrast, the *CLCbio* platform that has been developed as an independent variant detection platform for life-scientists with plug-ins that do not require computational coding to define variants [22]. In this report, we benchmark both *CLCbio* and *Varscan2* as two independent variant detection platforms to define the overlap in their mutation detection and define their dual utility in creating a mutant genomic reference database for optimizing mutant peptide detection using mass spectrometry (Fig. 1).

Exome Sequencing of DNA derived from A375 cells was performed using Agilent V5+UTR Exome Capture Kit and 100 bp paired-end reads were acquired using a coverage of 100x. Paired fastq files were imported into the *CLC Biomedical Genomics Workbench 3*. Adaptor sequences and bases with low quality were trimmed, DNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through dbSNP databases to remove “common” germline variants. A total of 120,325 DNA variants were detected at a frequency of 5% or higher at the threshold used (Supplementary Table 1A). Filtering this list of tumour specific variants to a frequency of 40% or higher in order to capture the most dominant mutant alleles in the cell population, resulted in a total of 63,880 variant mutations detected (Supplementary Table 1B).

When the *CLCBio* variant calling platform was compared to *Varscan2* using the hg19 reference database, as well as the more recently updated hg38 reference genome, then 85,793 shared variants were detected (Fig. 2A). The *CLCbio* platform detected more variants than *Varscan2* at the thresholds used; 36,065 variants were unique to *CLCbio* using the hg38 reference genome; 17,824 variants were unique to *CLCBio* using the hg19 reference genome, and 3,180 variants were unique to *Varscan2* using the hg38 reference genome (Fig. 2A). Because *CLCbio* generally identified more variants than *Varscan2*, we focused on using this platform to optimize mutant peptide detection by mass spectrometry (Fig. 1).

Using CLCbio RNA variant detection platform to develop a mutant peptide reference database from a human melanoma cancer cell line

Of the filtered *CLCbio* DNA variants detected (Supplementary Table 1B), 20,419 were synonymous mutations within exons, 41,993 variants resided out with exons, and the 1,468 were non-synonymous mutations. This number of non-synonymous mutations is within the expected range of a tumour like melanoma which has a relatively high number of single nucleotide variants (Supplementary Table 2). The 1,468 protein-coding variants were derived from 884 genes, including single-nucleotide polymorphisms, in frame-insertions, and in-frame deletions. This list is very conservative and could be expanded by including DNA sequencing reads below the 40% threshold level (Supplementary Table 1B). Representative *CLCbio* browser views summarizing DNA variants are represented in Fig. 2B and 2C.

We next used shotgun RNAseq data derived from A375 cells to identify expressed mutated genes under conditions in which more liberal variant calling in the DNA variant files could be tolerated. RNAseq from A375 cells was performed using human total RNA, depleted of ribosomal RNA, followed by random priming to generate cDNA. From this template paired-end Illumina HiSeq2500 was used to generate approximately 20 million reads. Paired fastq files (available upon request) from RNAseq reads were imported into the *CLC Biomedical Genomics Workbench 3*. The RNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through the A375 cancer genome sequence where at least 2 mutant DNA reads were identified, then dbSNPs were removed, to identify mutated and expressed genes. A total of 18,341 expressed non-synonymous mRNA variants were identified with a cutoff of 5% of the total RNA reads defined as mutated (Supplementary Table 3).

Although DNA variant calling would typically rule out the annotation of 1 mutant DNA sequencing read, we also filtered the fastq DNAseq files (Supplementary Table 2) against fastq RNAseq data (Supplementary Table 3) with a stringent cutoff requiring at least 40 mutated RNA sequencing reads (e.g. relatively highly expressed mutant alleles) and filtered against a more liberal DNA variant cutoff of at least 1 mutant genomic DNA sequencing read. This produced 5,980 RNA variants including synonymous, non-synonymous, and non-exonic mutations (Supplementary Table 4). Upon filtering for non-synonymous variants, a list was generated composed of 1,418 non-synonymous highly expressed RNA variants encoded by 976 mutant genes (Supplementary Table 5). We then determined the overlap of the 1,468 *CLCbio* derived DNA variants using the stringent DNA variant calling (Supplementary Table 2) to the 1,418 expressed mutated RNA variants identified using liberal DNA variant calling but requiring high levels of mutated mRNA reads; e.g. a highly expressed mutated gene (Supplementary Table 5).

The first thing to note is that the list of highly expressed mutant genes selected based on the numbers of RNA variant reads is highly divergent from the list of mutant genes selected based on

the number of DNA sequencing reads (Fig. 2D). Only 107 variant genes are shared in this subset. One example of an overlap between the DNA and RNA variant calling cutoff highlight the expressed mutant RNA derived from a mutant gene is the *gpatch4* gene (Fig. 2E). This gene is homozygous mutant (Supplementary Table 1A and B). A total of 877 out of 985 genes with expressed RNA variants are not present in the original DNA variant list (Supplementary Table 6). This produces a mutant gene expression rate of 107/774 or 12.1%. The true value will be higher than 12.1% since we removed RNA variants that exhibited lower than 40 reads. This mutant gene expression frequency is in lower range of ~30% previously published; previous studies have shown that 36% of validated somatic SNVs were observed in the transcriptome sequence when RNAseq data was compared with the genomes/exomes data in breast cancer [28] and similar proportions were also observed in a lymphoma study in which 137 somatic mutations were expressed in RNAseq, out of 329 total somatic mutations [29].

Mutated protein identification using the A375 DNA genomic reference database

We next aimed to define the extent to which the DNAseq or DNA+RNAseq reference databases could be used to identify mutated peptides by mass spectrometry. Proliferating A375 cells were lysed and protein was processed using the FASP (Filter-aided sample preparation) method [30]. Measured spectra were processed in ProteinPilot 4.5 search engine where a Swiss-prot and TrEMBL reference search database (Supplementary Table 7) was used (as described below) [31]. This produced a file of 949 wild-type proteins identified at FDR<1%. In order to determine whether any of these wild-type identified proteins are mutated, we next filtered the 1,468 non-synonymous DNA variant set (Supplementary Table 2) with the 949 detectable protein set (Supplementary Table 7) to generate a list of 42 potential mutant polypeptide sequences (Table 1). Only one high confidence mutated tryptic peptide covering a sequence of mutated protein from this group was identified. The peptide was derived from the ribosomal protein rpl14 (Fig. 3A and 3B). The mutant peptide spectrum and the fragmentation summary were exported from the ProteinPilot 4.5 search engine (Fig. 3C and 3F). The wt-rpl14 peptide covering the same position in protein sequence was also observed (Fig. 3D and 3E) which is consistent with the heterozygous mutation identified by DNA sequencing (Fig. 3A).

Limitations of mass spectrometry to identify mutated protein sequences.

There could be several reasons why the majority of these mutated proteins might not be detectable using mass spectrometry. First, as 18 out of the 42 proteins from this group have a homozygous gene mutation (Supplementary Table 1 and Table 1), these 18 expressed proteins presumably are mutated. However, there were no mutated tryptic peptides derived from any of these 18 proteins. This highlights a general difficulty in relying on mass spectrometry to confirm the expression of mutated proteins. The number of identified peptides covering any one protein is rarely “100%”. For example, the tryptic coverage of rpl14 highlights just this problem, as only two out of over 16 theoretical tryptic peptides (over 8 aminoacids of greater in length) could be detected (Fig. 3B). Serendipitously, one of these two peptides covered the mutant region (Fig. 3C and 3F). Other reasons for the absence of mutant peptide detection is that the mutant tryptic peptide could be relatively small (e.g. less than six amino acids in length) and therefore difficult to confidently match to corresponding MS/MS spectrum. For example, the protein LRCC59 has an in-frame triplet nucleotide insertion resulting in a Q amino acid insertion (Lys137_Pro138insGln) resulting in the tryptic penta-peptide QPFPK. It is difficult to identify this mutated peptide, unambiguously, as being derived from LRCC59 as there are several proteins in the human proteome that could yield this amino acid sequence after trypsinization at the sequence KQPFPK. The inability to detect such mutant tryptic peptides from any of the 18 proteins derived from homozygous mutant genes provides a measure of the false negative discovery frequency and

highlights the need to include different proteases in sample preparation or derive a larger pool of tryptic peptides from which to search for mutated peptides. Nevertheless, users of this A375 cell model as a resource might want to consider these 18 proteins (with no detectable mutant tryptic peptides) expressed from homozygous mutated genes to be *bona-fide* mutant proteins.

Mutated protein identification using mass spectrometry with the A375 mutant RNA reference database

As mutant peptides were not detectable to a high degree using the DNAseq-only files, we focused on using the RNAseq files stratified by 40 or more mutated mRNA reads derived from wt-p53 A375 cells (Supplementary Table 5) to create a mutant search database. We also next initiated an additional pipeline approach that was used to increase number of detectable proteins by pre-fractionating the peptides using an orthogonal LC method (Fig. 1). The orthogonal approach incorporated a reverse phase high pH acetonitrile gradient generating ten fractions that were infused for peptide identification using a TripleTOF 5600+ mass spectrometer. Employing peptide pre-separation step increased the coverage of total wild-type peptides to over 35,000 (FDR<1%) with a coverage of over 4,500 wild-type proteins (FDR<1%) (Supplementary Table 8).

Integrating the RNAseq derived mutant search database (Supplementary table 8) to the 2D LC-MS/MS data increased the number of identified mutant peptides to 193 (Supplementary Table 9). Although the ProteinPilot 4.5 search engine determines the confidence of identified mutated peptides, we needed to manually inspect the spectra in each of these 193 proteins to produce a list of mutated peptides that were identified based on y and b fragment ions covering the mutated amino acid in the mutated peptide sequence. Applying this procedure excludes the possibility that we identified wild-type peptides and narrowed the list down to 60 verified mutated proteins (Fig. 4).

Methodologies for validation of mutated peptides in the wt-p53 A375 cell model.

Mutated peptide identification in A375 fractions was further examined using targeted mass spectrometry in pseudo-selected reaction monitoring mode (pseudo-SRM) on TripleTOF 5600+ mass spectrometer. Ten isotope labeled mutant tryptic peptides were acquired (Fig. 5B) to validate ten of the 60 mutant tryptic peptides identified in data-dependent mode (Fig. 5A). We optimized isotope labelled peptide spike-in into A375 lysates (Fig. 5C and 5D) to yield optimal product ion intensity. Comparing the retention times and product ion intensity patterns in product ion chromatograms of naturally occurring intrinsic peptide and isotope labelled peptide enabled us to determine whether the naturally occurring peptide is present. All ten of the mutant peptides were successfully validated by this methodology.

As an example, verification of two mutated peptides is shown in Fig. 6A and 6B. Fragmentation evidences described in Fig. 6A and 6B show all possible product ions that represent two selected high confidence mutated peptides (peptide confidence > 99%) identified in 2D LC-MS/MS data (VSGSPEQAVEENLSSYFLDR and IIPTVLMTEDIK peptides). Only underlined product ion masses represent identified product ions that confirm the presence of amino acid mutation in the peptide sequence. The fragmentation evidence of the VSGSPEQAVEENLSSYFLDR mutated peptide highlights 11 product ions confirming the S to F mutation, whilst the fragmentation evidence of the IIPTVLMTEDIK mutated peptide shows only 3 product ions confirming the A to P mutation**.** The fragmentation evidences in Fig. 6A and 6B clearly show that the probability of an amino acid mutation is not reflected in the peptide confidence determined in ProteinPilot 4.5 software. Therefore, it is important to evaluate the spectra/fragmentation evidence to accurately define mutant status. An example SRM validation of a mutant peptide (SIITYVSSLYDTMPR) with heavy

isotope labelled reference peptide and its light naturally occurring variant is shown in Fig. 6C. All results from SRM validation of mutant peptides using SRM are summarized in Supplementary Table 10 and Fig. 6D. Taken together Fig. 6 summarizes all steps, that were taken to validate any of the selected mutated peptides.

Mutated proteomics: Creating a p53-null A375 melanoma cell line as an isogenic model system to define p53-dependence in mutated cancer genome protein expression patterns.
We finally aimed to use this optimized DNA and RNAseq variant detection pipeline (Fig. 1) to ask a key biological question; how does inactivation of the tumour suppressor p53 gene impact on the mutant proteome landscape? The human melanoma A375 cell line has classically been used as a model to study regulation of wt-p53 activity in response to DNA damage [26] [27]. We thus focused on using CRISP-R gene editing to develop an A375 cell line with an isogenic p53-null status. The guide RNA encoding a targeting sequence near the ARG175 codon in the p53 gene (Fig. 7A) was transfected into cells, single cells were isolated following cell sorting, and individual clones were selected based on absence of p53 induction after X-irradiation (data not shown). Nine independent p53-edited clones were obtained (data not shown). One A375 p53-null cell clone was taken forward with the edits as indicated in Fig. 7B and 7D. The sequences across the breakpoint in edited p53 alleles (Fig. 7) results in a stop codon (Fig. 7C and 7E). Immunoblotting of lysates using the N-terminal epitope antibody DO-1 confirmed that the A375 p53-null cells do not express p53 protein, nor do the HCT116 p53-null cells (Fig 7F). In addition, MHC Class I protein (HLA-B allele) was also determined to be elevated in the p53-null cells (Fig 7G). We also observed elevated HLA-A and HLA-C alleles in p53-null cells (data not shown), further highlighting the utility of this p53 –null cell as a tool to study in the future how p53 status impacts upon mutated protein expression as well as mutant peptide presentation by the MHC Class I system.

SWATH (Sequential Windowed Acquisition of all Theoretical Mass Spectra) was used as a complementary methodology to further determine the presence of target mutated peptides and to evaluate an effect of p53 inactivation on mutated protein levels. We first set-up large scale SWATH quantitation on TripleTOF 5600+ mass spectrometer and focused towards two mutated proteins that were previously verified and validated in wt-p53 A375 cells; PYGB and PLEC (Fig. 8). We extracted product ion chromatograms corresponding to these mutated peptides using a product ions m/z's included in a spectral library derived from data-dependent measurement of same sample. We found high correlation between product ion intensity pattern in extracted product ion chromatograms and the product ion intensities in spectral library for both peptides (Fig. 8A, 8B, 8D, 8E). Following quantification of the PYGB mutant peptide using three technical replicates obtained from A375 p53-null cells and A375 p53 wild-type cells shows an up-regulation of the PYGB (LIINLVTSIGDVVNHDPVVGDR) mutated tryptic peptide in A375 null cells. However, the quantification we observed based on the set of automatically intensity based selected product ions does not specifically represent the mutated peptide, and as such we cannot rule out the possibility that the quantitative data are derived from wild-type tryptic peptide. Therefore, as a robust SWATH pipeline we would again recommend to carefully select product ions referring uniquely to the mutation in a peptide sequence. In case of PYGB mutated peptide (LIINLVTSIGDVVNHDPVVGDR) it would be ions encompassing b4 – b21 and y19 – y21. The set of automatically selected product ions from spectral library contains y6, y16, y17, y9 (Fig. 8A). Therefore, SWATH quantitation of peptide LIINLVTSIGDVVNHDPVVGDR does not uniquely refer to the mutated peptide form and could be biased by changes in wild-type form.

We next extracted quantitative SWATH data for mutated peptide from PLEC (SIITYVSSLYDTMPR) that was also successfully verified and validated (Fig. 8D). In this example, we extracted

quantitative data referring to product ions that uniquely involve mutated amino acid in the product ion sequence. The mutated peptide SIITYVSSLYDTMPR must be quantitated by any product ions encompassing b12 – b15 and y4 – y15. The set of automatically selected product ions from spectral library for quantitation of SIITYVSSLYDTMPR 3+ contains y4, y5, y6, y7 (Fig. 8D). Therefore, quantitation of the mutated peptide refers uniquely to mutated peptide form. The corresponding spectral library evidence and product ion intensity rank in product ion chromatogram are highlighted (Fig. 8D and 8E) along with the quantitation of mutated peptides in three technical replicates of A375 p53-null cells and A375 p53 wild-type cells (Fig. 8F). As a control, SWATH quantitation of a control peptide VAPEEHPVLLTEAPLNPK (ACTB) was used to evaluate quantitative differences as a consequence of deleting the p53 gene in A375 cell line, differences in sample loading, and the possible effects of sample preparation on MS analysis (Fig. 8G).

Differences in mutant protein expression in wt-p53 and p53-null cells.
We finally determined the major differences in mutant peptides enriched between wt-p53 and p53-null cells using *CLCBio* and *Varscan* variant detection platforms. Parameters were set requiring at least 10 mutant mRNA variants and 1 mutant DNA variant. Using *CLCbio*, there were 190 mutant peptides detected enriched in wt-p53 cells, with 187 mutated peptides enriched in the p53-null cells (Fig. 9A). The largest number of mutant peptides were detected using *CLCbio* when compared to Varscan (Fig. 9B and 9C). Using STRING protein annotation, we evaluated the dominating mutant protein networks defined (Fig. 10 and 11).

In cells containing wt-p53 there were dominating peptides that could be used to highlight mutated protein networks centered on CDC5L (Fig. 11, center). The CDC5L network in turn is linked to a mutant protein splicing network composed of SRSF7, SF35B, SLU7, CSTF3, and YBX1. In turn these networks also connect to mutant proteins in DNA repair pathways including POLR1b, ERCC2, FANCD2, XRCC6, and others (Fig. 10). By comparison, the STRING analysis in p53-null cells highlighted a different dominating mutant proteome. This consisted of a ribosomal mutant protein network including RPL5 and RPL26L (Fig. 11). This node was in turn connected to a mutant ubiquitin protein node including USP7, UBE2V2, USP9X, and MRE11. An independent node composed of mutant proteins was present in kinase signalling including EGF, YES1, PPP2CB, CSK, and PPP2R5C (Fig. 11). Together, these data highlight that inactivation of the p53 gene can switch the expression of distinct mutant protein signalling nodes encoded by a cancer genome. This in part, can shed light on how tumour cells adapt to gene mutation by changing the expression of mutant proteins that comprise specific signalling pathways.

DISCUSSION

The identification of mutated proteins in human cancer cells can assist in defining expressed, mutated oncogenic signalling landscapes as well as facilitating the identification of potential neoantigen vaccine ligands. Most often, proteomics studies using cancer cell lines uses a normal reference proteome and under-estimates mutated protein signalling functions. We report on an optimized pipeline for identifying tumour variants in A375 cells using; (i) the *CLCBio* integrated DNA and RNA variant calling platform; (ii) the incorporation of RNAseq to stratify highly expressed variants; (iii) the use of 2D LC-MS/MS to identify potentially mutated peptide sequences in a tumor cell line; and (iv) the use of manual spectral annotation and SRM to estimate a false discovery rate of mutant peptides using LC-MS/MS. Our pipeline has identified high confidence list of protein mutations in the A375 cell line and its p53-null derivative (Supplementary Tables 13 and 14) by stratifying genetic mutations based on high levels of mutant mRNA expression and using

shotgun mass spectrometry to identify mutated peptides. This underestimates the extent of mutant protein expression in the A375 cell because of; (i) incomplete tryptic coverage of any one given protein; (ii) we only included very abundant mutant mRNA species (with relatively high numbers of reads; i.e. >40) to create a reference database of relatively highly expressed mutant genes; and (iii) stringent manual annotation of mutated peptides identified in the mass spectra eliminates some likely mutant peptides from the dataset.

We focused this study on using a human melanoma has emerged as a cancer type with one of the highest rates of single-nucleotide variation in a cancer genome and this tumour type is expected to form important models to define mutated protein expression networks [32]. As a result, melanoma patients can benefit from cancer-specific immuno therapeutics that exploit this high rate of mutated protein production [33]. Understanding how cancer associated genes impact on steady-state mutant protein expression levels and ultimately neoantigen assembly into the MHC Class I pathway has thus recently gained more relevance. Thus, it is important to begin to develop isogenic cancer cell models with specific cancer gene mutations to accelerate our understanding of how mutant protein production and re-wired mutated signalling proteins are enhanced in cancer cells. This could facilitate developing new therapeutics that exploit mutant protein or mutated pathway expression. We focused here on applying these methods to the question of how p53 gene inactivation can impact on mutant protein expression in an isogenic melanoma cell model.

P53 protein has been termed the guardian of the genome [34]. The p53 gene is one of the most frequently mutated in the vast majority of cancers [35] leading to loss of wt-function and enhanced genome instability [36]. It is therefore interesting that the majority of human melanomas retain wt-p53 alleles [32], possibly because the selections pressures driving the survival of cells with a mutant p53 gene is reduced due to frequent inactivation of the p53 regulatory *cdkn2a* locus in melanoma. Oesophageal adenocarcinoma also has very high rate of single nucleotide variation [37] but also a high rate of both p53 gene mutation and *cdkn2a* mutation [38]. These data might suggest there could be two distinct pathways that drive a high rate of single nucleotide variation in a cancer genome via attenuation of wt-p53 function (as in melanoma) or mutation in the p53 gene (as in oesophageal adenocarcinoma). We focus in this report on generating an isogenic wt and p53-null cell panel using CRISP-R/CAS9 gene editing methodology to define how loss of p53 can impact on a mutant proteomic landscape. Integrating genomics, RNA expression, and mass spectrometric data produces a baseline mutant protein library in melanoma that can be used as a community resource to facilitate interrogation of signal transduction pathways in this model. A similar cancer cell model has been developed in oesophageal adenocarcinoma that has the features of high rate of single nucleotide variation, but which has mutations in the p53 tumour suppressor gene [39]. Using our optimized pipelines, we define specific sets of mutant proteins that are differentially expressed as a result of p53 gene status. These data indicate that loss of a tumour suppressor gene such a p53 can begin to switch expression of the mutated protein landscape in a tumour cell (Fig. 10 and 11).

One impact of this pipeline will be in the future identification of tumour-specific mutated proteins that can be processed by the MHC Class I pathway; neoantigens. Within this list of mutant proteins detected in the A375 cell line, we were also able to identify some trimmed peptides that have a predicted high affinity for MHC class I peptides as defined using netMHC 4.0 [40], based on our isotyping that A375 cells have the MHC Class I alleles, HLA-A*-3:01 and HLA-B*07:02. One of these includes the gene HIST3H2A, which has a V108L mutation. The predicted affinity of the wt-

10mer LPNIQAVLL for HLA-B*07:02 is 518.5 nM and the predicted affinity of the mutant-10mer LPNIQAVVL for HLA-B*07:02 is 87.8 nM. Additionally, SRP14, with a P124A mutation resulted in a mutant 10mer APAAAATAA peptide with a predicted affinity of 17.9 nM for HLA-A*03:01 whilst the wt-peptide APAAAATAP peptide with a predicted affinity of 641.5 nM for HLA-A*03:01. Such datasets will provide neoantigen models to study MHC Class I peptide flux in this A375 cell line.

A second impact of this pipeline will be on mutant proteomics studies. We have used a gene editing tool (CRISP-R) to begin to ask a biological question on how loss of p53 can impact on the mutant protein landscape. Pathway annotation using STRING provides evidence for mutated protein expression in pathways including adhesion, ubiquitination, metabolism, and DNA repair. Mutant protein expression in a cancer cell line provides another way of thinking about 'proteomics' compared to its usual application which is examining protein expression in a cancer cell line using a 'normal' reference protein database. When screening the mass spectral data from A375 cells and the p53-null A375 cells against the mutant genomic reference databases, we identified a relatively high overlap in the proteins identified between the two cell lines (Fig. 9A). This might not be unexpected since p53 is stress activated. On the other hand, the difference in the proteome between the two cell lines (Fig. 9A) also highlights the fact that inactivation of the p53 gene did give rise to spontaneous changes in some mutated proteins without further selection pressures. Future studies could examine how loss of p53 in this isogenic model impacts on radiation or GAS/STING-dependent mutant protein signalling. Together, our study produces a well annotated mutant proteome of this A375 cell line model, coupled to the application of an integrated variant detection platform (*CLCbio*), that can be used to detect high confident mutant proteins cells. This can facilitate the use of this isogenic model as a resource to identify the p53-dependence on the mutant proteome and normal proteome landscape.

Acknowledgement: The work was partially supported by the European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868) and MH CZ - DRO (MMCI, 00209805), JF, PM and FZK were supported by AZV NV18-03-00339; the BBSRC RASOR consortium (BB/C511599/1; United Kingdom); the Wellcome Trust (grant number 094417/Z/10/Z); The International Centre for Cancer Vaccine Science project carried out within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund; the National Natural Science Foundation of China (Grant. 81701883) and China Scholarship Council.

Materials and Methods

P53 knockout using the CRISPR/CAS9 system.

The p53-specific gRNA sequence was 5'-CTGAGCAGCGCTCATGGTGGNGG-3', which was designed by Applied StemCell, Inc. The gRNA was cloned into pBT-U6-CAS9-2A-GFP expression vector from Applied StemCell, Inc. The P53 knockout in A375 cell line was performed as described before with minor alterations [41]. Briefly, $3x10^5$/well A375 cells [27] were seeded in 6-wells plates. 24 hours later, cells were transfected with pBT-U6-CAS9-2A-GFP expression vector using Attractene Transfection Reagent (QIAGEN, UK). 48 hours later, mutations were tested using a Surveyor Mutation Detection Kit (Integrated DNA Technologies, USA) and GFP positive cells were sorted and collected by BD FACSCanto II (BD Bioscience, USA). GFP positive cells were seeded in 96 wells plates 1 cell/well for colony formation. After 2 weeks, all colonies were collected and tested by western blot using the in-house developed DO-1 monoclonal antibody to demonstrate loss of p53 protein (Fig 7F) and sequencing to validate p53 gene editing in the A375 cell lines. The HLA-B antibody was from Thermofisher (PA5-35345). The loading control for immunoblotting (Fig 7F) was an in-house antibody developed to PCNA. The antibody used as a loading control (Fig 7G) was

mouse monoclonal anti-β-actin (Sigma). The HCT116 wt and p53-null cells were a gift of Dr. B Vogelstein (Johns Hopkins University, USA).

CLCbio variant calling.

Next generation DNA sequencing of the A375 wt-p53 containing melanoma cell line was performed in order to annotate its genome. Exome Sequencing was performed using Agilent V5+UTR Exome Capture Kit (75Mb) and 100 bp paired-end reads were acquired using a coverage of 100x. Paired fastq files from the A375 cell line (available upon request) from DNA-exome libraries were imported into the *CLC Biomedical Genomics Workbench 3*. Adaptor sequences and bases with low quality were trimmed, DNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through dbSNP databases to remove "common" germline variants. A total of 120,325 tumour specific variants were detected at a frequency of 5% or higher when the threshold was set at calling variants detected in at least two variant DNA sequencing reads in the exome data from A375 cells (Supplementary Table 1A). RNAseq from A375 cells was performed using human total RNA, depleted of ribosomal RNA, followed by random priming to generate cDNA. From this template paired-end Illumina HiSeq2500 was used to generate approximately 20 million reads. Paired fastq files (available upon request) from RNAseq reads were imported into the *CLC Biomedical Genomics Workbench 3*. The RNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through the A375 cancer genome sequence where at least 2 mutant DNA reads were identified, then dbSNPs were removed, to identify mutated and expressed genes. A total of 18,341 expressed non-synonymous mRNA variants were identified with a cutoff of 5% of the total RNA reads defined as mutated (Supplementary Table 3).

Varscan2 variant calling.

The same paired fastq files from the A375 cell line (as used for *CLCbio* analysis, above) from DNA-exome libraries were used for following analysis pipeline. Variant calling was performed using Varscan, adaptor sequences and bases with low quality (QualityScore < 30) were trimmed by FASTX-Toolkit version 0.0.14 (Retrieved from http://hannonlab.cshl.edu /fastx_toolkit), DNA sequencing reads were mapped to the human reference genome hg19 and hg38 (by TopHat2 [42]) (See Fig. 1), and results were imported into VarScan2 [43]. Sequences were filtered through dbSNP databases to remove "common" germline variants. A total of 101,072 tumor specific variants (coverage 1 DNA mutation) and 10,545 non-synonymous mutations were detected using hg38 (coverage 1 DNA mutation; See Supplementary Table 11). A total of 17,822 tumor specific variants were obtained from RNA seq (using a minimal cutoff of 10 mutated RNA reads) and 2,662 were classified as a non-synonymous RNA mutation (using a minimal cutoff of 10 mutated RNA reads). A total of 5,590 tumor specific variants were obtained from RNA seq (using a minimal cutoff of 40 mutated RNA reads, data not shown) and 1,007 were classified as non-synonymous RNA mutations (using a minimal cutoff of 40 mutated RNA reads). The number of 2,461 tumor specific variants from the wt-p53 A375 cell line (1 DNA mutation and at least 10 mutant RNA reads) and 945 tumor specific variants from the p53-null cell line (1 DNA mutation and at least 10 mutant RNA reads), were used as the input file for reference database to identify mutant peptides using mass spectrometry Supplementary Table 12. A comparison of the DNA mutations detected using Varscan and CLCbio are summarized in Fig. 2 and the number of mutated peptides detected, using either Varscan or CLCbio driven analysis are summarized in Fig. 9.

Peptide sample preparation for MS.

Cells were plated and grown on five 10 cm Petri dishes to 80% confluence. Cells were harvested into lysis buffer composed from 8 M urea in 0.1 M Tris/HCl pH 8.5 (urea buffer). Cell lysis was

facilitated using needle sonication in three 4 second cycles and protein concentration was determined using RC-DC assay (Bio-Rad, USA). An aliquot corresponding to 200 µg of protein was digested to peptides using Filter-aided sample preparation (FASP) [30]. Briefly, cell lysate was mixed with 200 µl of urea buffer and added to centrifugation filter unit Vivacon 500 with 10 kDa cut-off (Sartorius Stedim Biotech, Germany) followed by centrifugation (15000 g/ 20 min/ RT). 16.7 mM TCEP in urea buffer was added to filter unit to reduce disulphide bridges in protein. Reduction was done on thermomixer (600 rpm/ 30 min/ 37˚C) followed by centrifugation (15000 g/ 20 min/ RT). Sample alkylation was performed in the darkness for 20 min at RT with 300 mM Iodoacetamide in urea buffer followed by centrifugation (15000 g/ 20 min/ RT). Buffer was exchanged to 100 mM NH_4HCO_3 using 3 washes to enable efficient trypsin digestion of samples. Proteins were digested in 100 µl of 50 mM NH_4HCO_3 buffer where 1 µg of trypsin was added per each 33 µg of protein to be digested. The samples were mixed with digestion buffer at 600 rpm on thermomixer for 1 min before incubation for 18 h in a wet chamber, 37˚C. Peptides were eluted from the filter by centrifugation (15000 g/ 20 min/ RT). To increase peptide recovery 0.5 M NaCl was added followed by centrifugation (15000 g/ 20 min/ RT). Peptide samples for direct MS analysis were desalted on Micro SpinColumns C18 (Harvard Apparatus, USA). First, C18 columns were conditioned twice with 100% acetonitrile (AcN)/ 0.1% formic acid (FA), centrifuged at (120 g/ 2 min/ RT), and washed with 0.1% FA followed by centrifugation (200 g/ 2 min/ RT). The columns were hydrated for 15 min in 0.1% FA, centrifuged (200 g/ 2 min/ RT). Samples were loaded to columns and centrifuged (550 g/ 2 min/ RT). Next, the columns were washed thrice with 0.1% FA. Peptide elution was done using 50% AcN/ 0.1% FA in water, 80% AcN/ 0.1% FA in water and 100% AcN/ 0.1% FA. The peptide eluates were evaporated using a SpeedVac and dissolved in 5% AcN/ 0.05% trifluoroacetic acid (TFA) in water. Concentration of peptides was determined in each sample on NanoDrop 2000 (Thermo Scientific, USA) at 220 nm and 280 nm prior MS analysis.

RP-basic fractionation using spin columns.

FASP digested peptide samples for *RP-basic fractionation* were separated in basic pH (pH 10) on Macro SpinColumns C18 (Harvard Apparatus, USA) packed with C18 sorbent. First, mobile phase A composed from 10 mM ammonium formate (AF) in water pH 10 and mobile phase B composed from 10 mM AF in 90% AcN pH 10 were prepared [44]. C18 columns were conditioned twice with mobile phase A followed by centrifugation at (200 g/ 2 min/ RT). Columns were then washed with mobile phase B followed by centrifugation (300 g/ 2 min/ RT). Columns were hydrated for 15 min using mobile phase A, centrifuged (300 g/ 2 min/ RT). Peptide samples were added to hydrated columns and centrifuged (650 g/ 2 min/ RT). Next, the columns were washed thrice with mobile phase A. Peptide separation into 11 fractions was done using a step gradient composed from 5% B + 95% A, 9% B + 91% A, 13% B + 87% A, 17% B + 83% A, 21% B + 79% A, 25% B + 75% A, 35% B + 65% A, 50% B + 50% A, 80% B + 20% A and 100% B. Ten fractions were evaporated using a SpeedVac concentrator. Each fraction was dissolved in 100 µl of 50% methanol and then evaporated in SpeedVac concentrator. This step was repeated three times to get rid of volatile salts. Dried samples were dissolved in 20 µl of 5% AcN / 0.05% TFA. Concentration of peptides was determined in each sample on NanoDrop 2000 (Thermo Scientific, USA) at 220 nm and 280 nm prior MS analysis.

Spiking sample with reference heavy peptides.

JPT synthetic heavy reference peptides (JPT, Germany) derived from 10 mutant peptide candidates were ordered with isotopically labelled amino acids. A content of the vial containing heavy reference peptide was dissolved in 5% AcN + 0.05% TFA in water to prepare 1.43 nmol/µl stock solution. Next, LOD of each peptide was determined (data not shown). A pool representing all 10 reference peptides was prepared and spiked into each peptide fraction (A375 peptide

concentration and injection volume was set based on crude Nanodrop measurement A_{220} and A_{280}) in order to load onto the column an identical amount of reference peptide corresponding to at least 10 times LOD. Three µl of this mixture were injected onto nano-LC-MS/MS to perform pseudo-SRM analyses in analytical triplicates.

LC setup for analysis of fractions.

Eksigent Ekspert nanoLC 400 system (AB-SCIEX, USA) nano-LC system was used for peptide concentration and separation. Peptides were concentrated and desalted on a cartridge trap column (300 µm i.d. × 5 mm) packed with C18 PepMap100 sorbent with 5 µm particle size (Thermo Fisher Scientific, MA, USA). Peptides were 10 min washed using 0.05% TFA in 5% AcN in water. Separation was performed on a 25 cm fused-silica emitter column with 75 µm inner diameter (New Objective, USA), packed in-house with ProntoSIL C18 AQ 3 µm beads (Bischoff Analysentechnik GmbH, Germany). LC solvents were composed from 0.1% FA in water (solvent A) and 0.1% FA in AcN (solvent B). Sample was eluted in a 120 min gradient starting at 5% B up to 40% B with a flow rate 300 nL/min in DDA and SWATH experiments while 61 min gradient starting at 5% B up to 40% B with a flow rate 300 nL/min was used in pseudo-SRM experiments. Peptides eluting from column were ionized in nano-electrospray and entered mass spectrometer.

Shotgun MS/MS (DDA).

Mass spectrometer TripleTOF 5600+ (AB-SCIEX, Toronto, Canada) operated in data dependent mode. Each cycle was accompanied with fragmentation of top 20 most intense precursor ions. Exclusion time was set to 12 seconds. Minimum precursor ion intensity was set to 50 cps and 100 milliseconds accumulation time per precursor. Shotgun were searched using ProteinPilot 4.5 (AB-SCIEX, Canada) against custom built human mutant proteome reference database derived from human reference database (Uniprot+Swissprot 2016_2) and against wild-type human reference proteome database (Uniprot+Swissprot 2016_2). Search parameters were set as follows: trypsin protease, carbamidomethyl (C) (fixed). Protein FDR was determined by searching MS/MS data against decoy databases.

Building-up custom search library with mutant protein sequences.

MS/MS data from fractions were inspected for presence of MS/MS spectra corresponding to genes identified as high confidence RNA and DNA variant hits. Amino acid sequences corresponding to these genes were listed in a mutant search database where a mutant position in a protein was inserted based on genomic data. Mutant search database was assembled in FASTA format. Corresponding wild-type FASTA sequences of protein forms were downloaded from the human (2016_2) reference database. A custom mutant FASTA file containing mutant and WT forms of proteins was created in a text editor and subsequently imported into ProteinPilot 4.5. Quality of mutant IDs was inspected in ProteinPilot 4.5. (AB-SCIEX, USA), focusing mainly to a peptide covering mutation position in a sequence. Spectral evidence of high and mid confidence mutant peptides (peptide confidence > 95% and between > 50% and < 95%) was manually inspected. In mutant peptides we focused to corresponding product ions directly proving a shift in mass as a consequence of mutation. A protein referred as “mutant” had covered potential mutation position by high or mid confidence peptides bearing substitution, insertion, deletion of amino acid in its sequences. On the other hand, a protein referred as “wild-type” has a place of potential mutation covered by wild-type sequence of a high or mid confidence peptide. Identification of mutant or wild-type form was accomplished by Uniprot BLAST of peptide sequence against Human reference proteome Uniprot 2016_2 to prove its proteotypicity.

SWATH-MS.

SWATH method for label free quantification was developed according to previously published methods [45] [46] [47]. TripleTOF 5600+ (AB-SCIEX, Canada) operated in high sensitivity positive polarity SWATH mode. Effective precursor range was selected from 400 amu up to 1200 amu and the cycle time was set to 3.5 seconds. An optimal SWATH width was 13 Da including 1 Da overlap resulting in a method with 67 SWATH windows. Accumulation time per SWATH was 50 milliseconds. Product ion range was scanned from 360 amu up to 1360 amu and rolling collision energy was used with collision energy spread (CES) of 15 mV.

Building-up wild-type spectral library and SWATH quantification of wild-type proteins.
PeakView software 1.2.0.3 (AB-SCIEX, Canada) was used to index 1045 proteins FDR 1% in a spectral library. Retention time window was set to 2.5 min to the left and 2.5 min from expected retention time. Protein quantitation was based on extracting peak areas for four peptides per protein and 6 product ions per each peptide. Extracted quantitative data from three technical replicates were statistically evaluated in MarkerView 1.2.1.1 (AB-SCIEX, Canada). Pairwise T-test was performed to determine protein fold changes and *P* values of fold change for all proteins listed in spectral library.

Building up a mutant spectral library and SWATH quantification of mutant proteins.
Results from mutant database MS/MS search were imported into Skyline-daily (64-bit) software version 2.6.1.6899 (MacCoss Lab, WA, USA) where the mutant spectral library was generated. FASTA sequences of 10 candidate mutant proteins listed in Supplementary Table 10 were imported into Skyline software. Results from mutant peptides are shown in Supplementary Table 10. An example of mutant peptides from RPL14 and PLEC is shown in Fig. 8. Peptide settings were as follows: digestion – trypsin, 1 missed cleavage was permitted, length of peptide was in range from 7 – 25 amino acids, 25 amino acids from N-terminus were excluded, carbamidomethyl was set as structural (fixed) modification. Transition settings were as follows: precursor charge was 2+, 3+ or 4+, fragment charge 1+, fragment series y or b, product ions, fragment ions from 3 up to last, including the N-terminal fragment to proline. After data extraction, start and end points of each peak in extracted product ion chromatogram were inspected manually. We checked co-elution of selected product ions, product ion ranks according to the spectral library, retention times across replicates, and data acquisition modes to confirm peak identity. Peak areas corresponding to mutant peptides were evaluated and visualized in Skyline software.

Pseudo-SRM analysis.
The TripleTOF 5600+ (AB-SCIEX, Canada) was operated in high sensitivity positive mode. Each cycle involved one TOF-MS scan with 250 ms accumulation time and 38 product ion scans of 70 milliseconds accumulation time per precursor. Total cycle time was 3.0 seconds. Product ions were scanned in a range from 100 to 1800 amu. Pseudo-SRM data were analysed in *Skyline*. Sequences of ten mutated peptide candidates were imported into the *Skyline* version 2.6.1.6899 (MacCoss Lab, WA, USA) with the following settings: Peptide settings were as follows: digestion – trypsin, 1 missed cleavages were permitted, length of peptide was in range from 7 – 25 amino acids, 0 amino acids from N-terminus were excluded, carbamidomethyl was set as structural (fixed) modification, heavy isotope peptide labeling 13C 15N on lysine, arginine and leucine was permitted. Transition settings were as follows: precursor charge was 2+, 3+, fragment charge 1+, fragment series y or b, product ions, fragment ions from 3 up to last, including the N-terminal fragment to proline. After data extraction, start and end points of each peak in extracted product ion chromatogram were inspected manually. We checked co-elution of selected product ions, product ion-rank according to the spectral library, retention times across replicates, and data acquisition modes to confirm peak identity. Peak areas corresponding to mutant peptides were evaluated and visualized in

Skyline software. dot-product (dotp) correlation for the ratio of the observed SRM peak intensities of a peptide in a specific biological matrix as correlation of observed SRM peak intensities of a peptide in a specific biological matrix versus the reference isotope labelled peptide were calculated. Peptides considered as high-quality validated hits showed good signal and dotp>0.9 and equal retention time to isotope labelled reference peptide.

FIGURE LEGENDS

Figure 1. Experimental plan for optimizing mutant protein detection in cell models. Two different variant calling platforms, *CLCBio* and *Varscan*, were used to identify DNA variants from the A375 cell line. Shotgun RNA seq was then performed on both parental and p53-null A375 cell lines to create a mutant expressed reference database for both cell lines. This mutant RNA database forms the reference from which mutant proteins will be identified using fractionation of the mutant tryptic peptides followed by mass spectrometry.

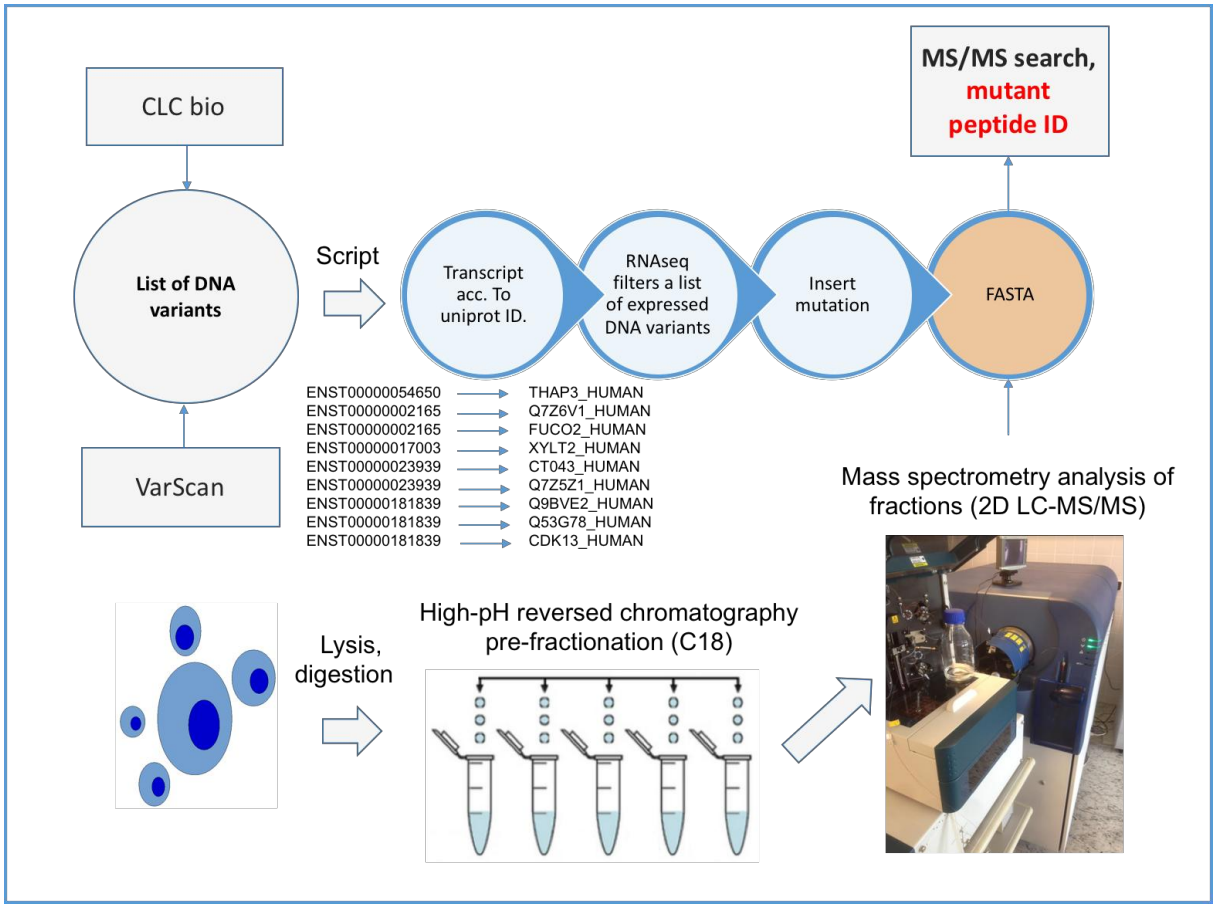


Figure 2. Representative variant calls from mutated genes. (A). Summary of the number of DNA mutations detected using the *CLCBio* and *Varscan* variant calling platforms. (B and C). Example *CLCbio* browser highlights mutated genes including a single nucleotide variant (MEIS1, chr2, 66691338) and an and in-frame insertion (SYNGR1, chr22, 39777822). The data are plotted by (i) chromosome position including hg19 reference genome (top, numbering); (ii) intron-exon boundaries. The thick line and thin line represent the exon and intron, respectively. Reference sets include: the blue line, that highlights ensemble v74 gene locus; the green line that represents ensemble v74 mRNA; and the yellow line that represents ensemble v74, CDS; (iii) highlights some of the DNA sequencing reads in blue aligned to hg19 with the DNA mutation variant (in A and B) highlighted by a color change and arrow**.** (D). *An analysis of shared DNA v RNA mutations in A375 cells*. A comparison of the overlap of variants detected in RNAseq and DNAseq filtered through distinct processes. The variants identified by DNAseq were filtered based on the presence of at least 4 mutant sequencing reads and at least a 50% frequency. A total of 1,468 variants were detected in 887 genes. The variants identified in mutated RNA were filtered by requiring at least

40 mutant RNA sequencing reads and at least 1 mutation in the DNAseq. This generated a different list of 1,418 mutated genes with highly expressed mutated RNA. Fusing the datasets produces a relatively small overlap of 107 common variants. The data suggest that the majority of mutated genes with a high confidence variant calling (774/887) are not highly expressed. *E. An example of one of the 107 shared mutated RNA and DNA CLCbio defined variants from the data filtered in Fig. 2D. GPATCH4 RNAseq* reads are highlighted as an example containing a frame-shift insertion mutation (AC) in DNA sequencing reads (Panel (I)) and in RNA sequencing reads (Panel (II)).

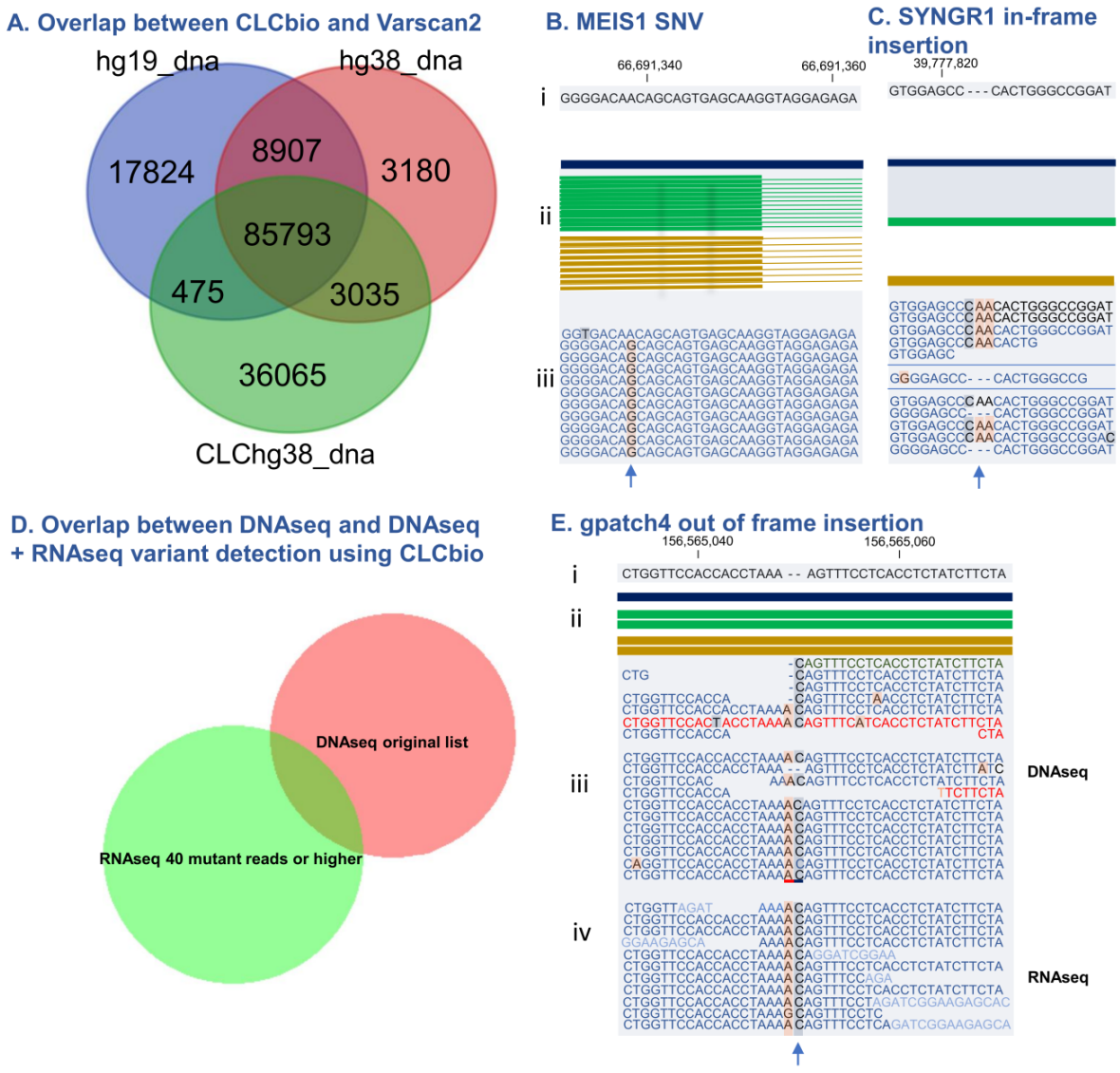
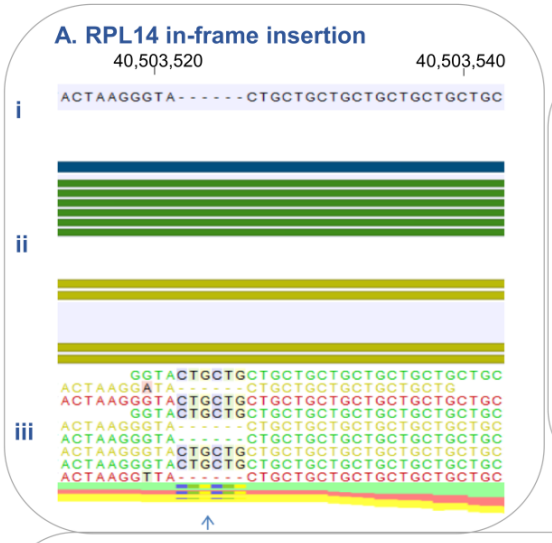


Figure 3. Identification of mutated rpl14 protein in A375 cells. The reference mutated DNA database was used to screen peptide spectra for evidence of mutant tryptic peptides. One mutant protein was detected by this method; rpl14. (A). The *rpl14* mutation is generated by an in-frame insertion of 6 bases encoding an AlaAla. The mutation is heterozygous as defined by the presence of 12 mutant reads and 11 wild-type reads (from Supplementary Table 1 and highlighted in the browser view by gaps in the sequence that matches the hg19 reference). The data are plotted by (i) chromosome position including hg19 reference genome (top, numbering); (ii) intron-exon boundaries. The thick line and thin line represent the exon and intron, respectively. Reference sets include: the blue line, that highlights ensemble v74 gene locus; the green line that represents ensemble v74 mRNA; and the yellow line that represents ensemble v74, CDS; (iii) highlights some of the DNA sequencing reads in blue aligned to hg19 with the DNA mutation variant highlighted by a color change and arrow. (B). The mutant protein tryptic coverage of rpl14 protein with the green highlighting high confident detection of two tryptic peptide sequences in (in green). (C and F). The mutant tryptic sequence containing 12 strings of alanine GTAAAAAAAAAAAK (mass 1156.60) has a 99% confidence of identification after fragmentation using ProteinPilot and there is no other such peptide in the human database. (D and E). The wild-type sequence containing 10 strings of alanine GTAAAAAAAAAAK (mass 957.5117) was also detected and has a 90% confidence of identification using ProteinPilot and there is no other such peptide in the human database. The existence of both tryptic peptides is consistent with the heterozygous nature of the tumour cell line (A).



B. Tryptic coverage of mutated RPL14 (ins AA)

MVFRRFVEVGRVAYVSFGPHAGKLVAIVDVIDQNRALVDP
CTQVRRQAMPFKCMQLTDFILKFPHSAHQKYVRQAWQK
ADINTKWAATRWAKKIEARERKAKMTDFDRFKVMKAKK
MRNRIIKNEVKKLQKAALLKASPKKAPGTKGTAAAAAAA
AAAAAAKVPAKKITAASKKAPAQKVPAQKATGQKAAPAP
KAQKGQKAPAQKAPAPKASGKKA

green = peptide confidence >95%
orange = peptide confidence <95% and >50%
red = peptide confidence < 50%
grey = no peptide coverage

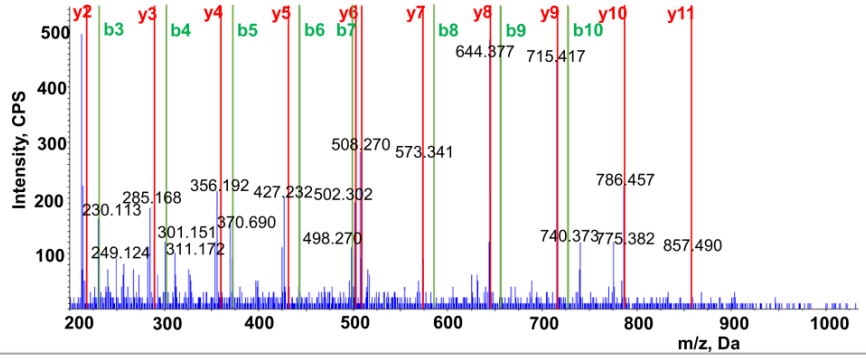
C. RPL14 (ins AA) GTAAAAAAAAAAAAK

Residue	Product Ion series b	b+2	y	y+2
1. G	58.0287	29.5180	1157.6273	579.3173
2. T	159.0764	80.0418	1100.6058	550.8066
3. A	230.1135	58.0287	999.5582	500.2827
4. A	301.1506	151.0790	928.5211	464.7642
5. A	372.1878	186.5975	857.4839	429.2456
6. A	443.2249	222.1161	786.4468	393.7271
7. A	514.2620	257.6346	715.4097	358.2085
8. A	585.2991	293.1532	644.3726	322.6899
9. A	656.3362	328.6717	573.7064	287.1714
10. A	727.3733	364.1903	502.2984	251.6528
11. A	798.4104	399.7089	431.2613	216.1343
12. A	869.4476	435.2274	360.2241	180.6157
13. A	940.4847	470.7460	289.1870	145.0972
14. A	1011.5218	506.2645	218.1499	109.5786
15. K	1139.6167	570.3120	147.1128	74.0600

D. RPL14 wild-type GTAAAAAAAAAAK

Residue	Product Ion series b	b+2	y	y+2
1. G	58.0287	29.5180	1015.5531	508.2802
2. T	159.0764	80.0418	958.5316	479.7694
3. A	230.1135	115.5604	857.4839	429.2456
4. A	301.1506	151.0790	786.4468	393.7271
5. A	372.1878	186.5975	715.4097	358.2058
6. A	443.2249	222.1161	644.3726	322.6899
7. A	514.2620	257.6346	573.3355	287.1714
8. A	585.2991	293.1532	502.2984	251.6528
9. A	656.3362	328.6717	431.2613	216.1343
10. A	727.3733	364.1903	360.2241	180.6157
11. A	798.4104	399.7089	289.1870	145.0972
12. A	869.4476	435.2274	218.1499	109.5786
13. K	997.5425	499.2749	147.1128	74.0600

E. Wild-type peptide GTAAAAAAAAAAK fragmentation spectrum



F. GTAAAAAAAAAAAAK (ins AA) fragmentation spectrum

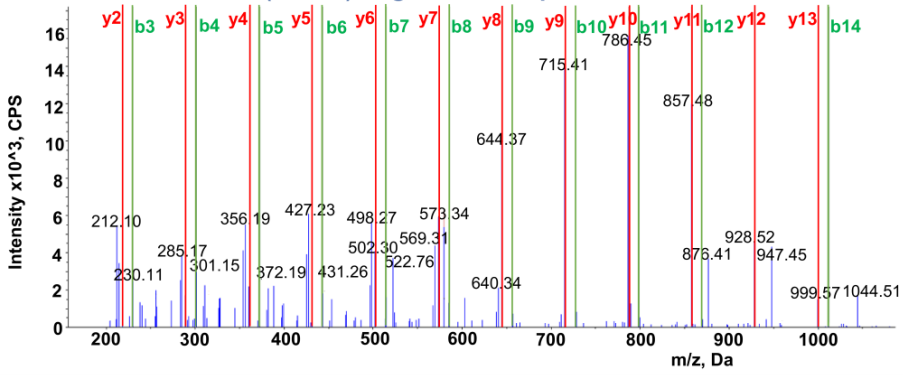


Figure 4. Optimization of mutant peptide identification using combined RNAseq and 2D LC-MS/MS datasets. A mutated reference search database file was based on RNAseq expression data (Supplementary Table 5) and then used to search 2D LC-MS/MS data leading to an increase in the number of peptides derived from potentially mutant proteins to 193 (Supplementary Table 9). The spectra in each of these were manually evaluated for the presence of y and b fragment ions that cover mutated amino acids characteristic by a m/z shift relative to product ions derived from the wild-type peptide. The data highlight successfully inspected mutated peptide candidates with an evidence of mutation in their peptide sequence. Tabulated are; Uniprot accession; mutation; mutated peptide sequence; peptide confidence; peptide charge state; peptide modifications and miscleavages; and product ions providing evidence of mutation (product ions that covered the mutant position in a peptide). The underlined peptides in the neoantigen sequence column highlights peptides that were subsequently validated using SRM.

Uniprot accession	Mutation	Mutated peptide sequence	Peptide confidence	Peptide charge state	Peptide modifications and misscleavages	Detected product ions proving evidence of mutation
H2A3	Val108Leu	LGRVTIAQGG**L**LPNIQAVLLPK	99	2+, 3+, 4+	cleaved L-L@N-term; missed R-V@3	b11, b12, b14, b15, b12 2+, b15 2+, b16 2+, b17 2+, b 19 2+, b20 2+, y11, y11 2+
SRP14	Pro124Ala	AAAAAAAAAPAAAATA**A**TTAATTAATAAQ	99	3+, 4+	-	y13, y14
CAN2	Asp22Glu	DREAAEGLGSHER	99	3+	missed R-E@2	y2, y3, y4, y5, y6, y7, y8
EZRI	Ile580Val	VDEFEA**L**	99	2+	-	b2, b3, b4, b5, b6
GSTO1	Ala140Asp	EDY**D**GLKEEFR	98	3+	missed K-E@7	b5, b7, y8, y8 2+, y9 2+, y10 2+
A0A0A0MTS2	Ile223Thr	TLAQLNPESSLFI**T**ASK	99	2+, 3+	-	y4, y5, y6, y7, y8, y9, y10, y11, y12, y13
TRAP1	Asp395Glu	LVSDGQALPEM**E**IHLQTNAEK	99	3+	-	y10, y11, y12, y13, y13 2+, y14 2+, y15 2+, y16 2+, y17 2+, y18 2+, y19 2+,
PHB	Arg43Leu	F**L**GVQDIVVGEGTHFLIPWVQKPIIFDCR	99	2+, 3+	Carbamidomethyl(C)@28	b2, b3, b4, b5, b6, b7
SPNXC	Leu68Val	TSPEEL**V**NDHAR	1	2+	missed R-E@12	y9
SUCB2	Thr396Ala	LEG**A**NVQEAQK	99	2+	-	b4, b5, b6, b7, b8, b9, b10, y8, y9, y10
TACC3	Ser190Phe	VSGSPEQAVEENLSSY**F**LDR	99	2+, 3+	-	y4, y5, y6, y7, y8, y9, y11
PLEC	Arg398Thr	SIITYVSSLYD**T**MPR	99	2+, 3+	-	y4, y5, y6, y7, y8, y9, y10, y11, y12, y13
MYH9	Ile1626Val	DLEAH**V**DSANK	99	2+, 3+	-	b6, b7, b8, b9, b10, y6, y7, y8, y9
RAD18	Arg302Gln	SAAEIV**Q**EIENIEK	99	2+, 3+	-	b8, b10, b12, b13, y8, y9, y10, y11, y12
PUR9	Thr116Ser	TVASPGV**S**VEEAVEQIDIGGVTLLR	99	3+, 4+	-	b8, b10, b11, b12, b13, b14, b15
B4DUC8	Val73Ile	NVDC**I**LLAR	99	2+	Carbamidomethyl(C)@4	b5, b6, b7, b8, y5, y6, y7, y8
ECHM	Thr75Ile	**I**FEEDPAVGAIVLTGGDK	99	2+, 3+	-	b2, b3, b4, b5, b6, b7, b8
PSD13	Asn13Ser	DVPGFLQQSQ**S**SGPGQPAVWHR	99	3+	-	y12, y13, y14, y12 2+, y14 2+, y15 2+, y16 2+, y17 2+, y18 2+, y19 2+, y20 2+
PDLI5	Ser492Asn	ILGEVI**N**ALK	99	2+	-	y4, y5, y6, y7, y8, y9, b8, b9
C1TC	Lys134Arg	DVDGLTSINAG**R**	99	2+	-	y4, y5, y6, y7, y8, y9, y10
K2C8	Arg369His	ASLEAAIADAEQ**H**GELAIK	99	2+, 3+	-	y7, y8, y9, y10, y11, y12, y13, y14, y15, y10 2+, y11 2+, y12 2+, y13 2+, y14 2+, y15 2+, y16 2+, y17 2+, y18 2+
GEMI4	Ala579Gly	FL**G**QILTAFPALR	99	2+, 3+	-	b3, b4, b5, b6, b7, b8, b9, y11, y12, y13 2+
PSB4	Ile234Thr	FQ**T**ATVTEK	99	2+	-	b3, b5, b6, b7, b9 2+, y7, y8
ANM3	Ser508Asn	VTVHK**N**K	84.8	2+	missed K-N@5	b6, y3, y4, y5, y6
A0A0A0MS30	Ala908Thr	LNGQQ**T**ALASQYR	98.8	2+	-	b10, y8, y9, y10, y13 2+
ESTD	Gly190Glu	AFS**E**YLGTDQSK	99	2+	-	b4, b5, b6, b7, b8, y9, y10, y12 2+
EMC1	Ser344Thr	**T**SSSEDGSMGSFSEK	99	2+	-	b3, b4, b5, b6, b7, b10 2+, b15 2+
AIP	Gln228Lys	EQPGSPEWIQLD**K**Q	98.2	2+	-	y2, y3, y4, y5, y6, y7, y8, y9, y10, y11, y12
RT27	Gly298Asp	EALDVL**D**AVLK	99	2+	-	b11 2+, y5, y6, y7, y8, y9, y10, y11 2+
UBQL4	Ile495Met	AMQALLQIQQGLQTLQTEAPGLVPSLGSFG**M**SR	95.5	4+	-	y4, y7, y8, y9, y10, y13, y14
SHCBP	Met21Thr	**T**GWAVEQELASLEK	99	2+	-	b3, b4, b5 2+
-	Ser259Gly	LQPLLNHLSHSYT**G**QDYSTQGNVGK	99	4+	-	b15 2+, b16 2+, b17 2+, b18 2+, y12, y14
RPC10	Ser24Ala	F**A**CNTCPYVHNITR	99	3+	Carbamidomethyl(C)@3, @6	b2, b3, b5, b12 2+
Q5SRN5	Glu176Val	WEAAH**V**AEQLR	64.1	3+	-	y6, y7, y8, y6 2+, y7 2+, y8 2+, y9 2+, y10 2+
CND1	Gln83Glu	SIDPGLKEDTL**E**FLIK	32	3+	missed K-E@7	b13 2+, b 15 2+, y5, y7, y8, y9, y10, y13 2+, y14 2+
PYGB	Lys622Asn	LII**N**LVTSIGDVVNHDPVVGDR	99	3+	-	b4
RD23B	Ala249Val	ESQAVVDPPQAASTG**V**PQSSAVAAAAATTTATT TTTSSGGHPLEFLR	99	4+, 5+	-	b16, b17, b33 2+, y33 2+
J3KQ32	Tyr274Cys	**C**DPGALVIPF	79.2	2+	Carbamidomethyl(C)@1, cleaved F-S@C-term	b2, b5, b6, b7, b8,
PPID	Leu302Ile	MSNWQGAIDSCLEALE**I**DPSNTK	99	3+	Carbamidomethyl(C)@11	y7, y8, y9, y10, y9 2+, y10 2+, y11 2+, y12 2+, y13 2+, y14 2+, y15 2+,
EFGM	Val215Ile	GI**I**DLIEER	99	2+	-	b3, b4, b5, b6, b7, y7, y7 2+, y9 2+
Q5SRN5	Asp185Glu	AYL**E**GTCVEWLR	99	2+	Carbamidomethyl(C)@7	b4, b5, b9, b11, b12 2+, y9, y10, y11
GRWD1	Arg8Pro	ESALEPG**P**VPEAPAGGPVHAVTVVTLLEK	98.5	3+	-	b9, b12, y22 2+, y24 2+, y25 2+
HEXA	Ile436Val	DFY**V**VEPLAFEGTPEQK	99	2+	-	b4, b5, b6, b4 2+
HEAT1	Asn1694Ser	NFGAENPDPFVPVL**S**TAVK	99	3+		y5, y6, y7, y8, y9, y10, y11, y7 2+, y8 2+
NEST	Val130Ala	AWLSSQ**A**AELER	90.2	2+	-	y6, y7, y8, y9, y10, b8
ADT2	Leu111Arg	YFAGNLASGGAAGATSLCFVYPLDFAR	99	3+	Carbamidomethyl(C)@18	b2, b3, b4, b5, b6, b7, b8, b10, b11, b12, b13, b14, b15, b16, b17, b9 2+, b10 2+, b11 2+, b13 2+, b14 2+, b15 2+, b17 2+, b18 2+, b20 2+
COPE	Thr117Ile	SVDVTN**I**TFLLMAASIYLHDQNPDAALR	99	3+, 4+	-	b10, b7 2+
EXOS7	Val274Leu	VLHASLQSV**L**HK	99	2+	-	b10, y7, y8, y9, y10, y10 2+, y12 2+
BIRC6	Val1332Leu	CAM**L**QFSEFHEK	1	2+, 3+	Carbamidomethyl(C)@1; Deamidated(Q)@5	b4, b5, y9, y12 2+
UBP24	Val2468Ala	NTFQLLHEILVIEDPIQ**A**ER	99	3+	-	y4, y6, y7, y8, y9, y10, y11, y6 2+, y10 2+
LGUL	Glu111Ala	ATLELTHNWGTEDD**A**TQSY	1	2+, 3+	cleaved Y-H@C-term	y19 2+
BRE1B	Gln615Arg	EREGPSLGPPPVASALSRADR	1	3+	missed R-E@2	b3, b5, b6, b7
SAM50	Ile345Val	FYLGGPTS**V**R	91	2+	-	y2, y3, y4, y5, y6, y7, y8, y9
PSB3	Met34Leu	FGIQAQ**L**VTTDFQK	99	2+	-	b7, b8, b11 b13, y9, y10, y11, y12, y13, y14 2+, y11 2+, y12 2+, y14 2+
C9JJ19	Leu109Phe	VRPDYTAQN**F**DHGK	99	3+	-	y5, y6, y7, y8, y9
IN35	Met128Val	VQVQPLELPMVTTIQVM**V**SSQLSGR	99	3+	-	y8, y9, y10, y11, y12, y13, y14
UBP24	Thr226Ile	NTFQLLHEILVIEDPIQ**A**ER	99	3+	-	y4, y6, y7, y8, y9, y10, y11, y6 2+, y10 2+
LAMC1	Leu888Pro	ACNCN**P**YGTMK	99	2+	Carbamidomethyl(C)@2; @4	y6, y7, y8, y9, y10 2+, y11 2+
RPL14	Ala157Ala-Ala ins	GTAAAAAAAAAA**AA**K	99	2+	-	b13, b14, y12, y13
HEAT1	His348Arg	LCNVPDLITILHGISETYDVSPLL**R**	99	3+	Carbamidomethyl(C)@2, cleaved H-L@N-term	y4, y6, y7, y11, y12, y13, y14

Figure 5. SRM validation of mutant peptide levels. *(A).* A List of identified mutated proteins and *(B).* corresponding peptides selected for validation using pseudo-selected reaction monitoring (pseudo-SRM). *(B).*highlights peptide sequences with green letters representing amino acids that are isotope labelled in the reference peptides. Red letters represent positions in peptide sequences that are mutated. *(C).* and *(D).* represent optimization of isotope labelled reference peptide QRVDEFEAL spiked-in into the A375 lysate. Extracted product ion chromatogram *(C).* shows intensity of SRM transitions from QRVDEFEAL mutated reference isotope labelled peptide at various concentrations, while D shows peak areas of corresponding product ions (y3 - y8). From *(C).* and *(D).* we determined that optimal spike in of QRVDEFEAL reference peptide is approximately 14.3 fmol which corresponds to a 10,000x dilution of stock reference peptide solution. This same procedure was carried out for all 10 heavy labelled peptides spiked into A375 lysates.

A. Mutated peptide chosen for p-SRM

Protein	Neopeptide sequence	Mass	Charge	Mutation
Keratin, type II cytoskeletal 8_KRT8	ASLEAAIADAEQHGELAIK	1935,9811	3+,2+	R to H
HLA class I histocompatibility antigen_HLA-A	WEAAHVAEQLR	1495,7169	2+	D to E
ADP/ATP translocase 2_SLC25A5	YFAGNLASGGAAGATSLCFVYPLDFAR	2795,3381	3+	L to R
Ezrin_EZR	QRVDEFEAL	1088,5101	2+, 2+	I to V
Ezrin_EZR	VDEFEAL	821,3779	2	I to V
Plectin_PLEC	SIITYVSSLYDTMPR	1744,8699	2+,3+	A to T2
Glycogen phosphorylase, brain form_PYGB	LIINLVTSIGDVVNHDPVVGDR	2344,2761	3+	K to N1
RPL14 protein_RPL14	GTAAAAAAAAAAAAK	1156,6073	2+	AA insertion
Transforming acidic coiled-coil-containing protein 3_TACC3	VSGSPEQAVEENLSSYFLDR	2226,0449	3+	S to F
PDZ and LIM domain protein 5_PDLIM5	ILGEVINALK	1068,6533	2+	S to N

B. Heavy mutated peptide chosen for p-SRM

Peptide Name	Sequence	Peptide Length	Stable Isotope-Labeled Residue
1	ASLEAAIADAEQHGELAIK	19	Lysine (K), +8Da
2	*WEAAHVAEQLR	11	Arginine (R), +10Da
3	*YFAGNLASGGAAGATSLCFVYPLDFAR	27	Arginine (R), +10Da
4	QRVDEFEAL	9	Arginine (R), +10Da
5	VDEFEAL	7	Leucine (L), +7Da
6	SIITYVSSLYDTMPR	15	Arginine (R), +10Da
7	LIINLVTSIGDVVNHDPVVGDR	22	Arginine (R), +10Da
8	GTAAAAAAAAAAAAK	15	Lysine (K), +8Da
9	VSGSPEQAVEENLSSYFLDR	20	Arginine (R), +10Da
10	ILGEVINALK	10	Lysine (K), +8Da

* new cleavage site

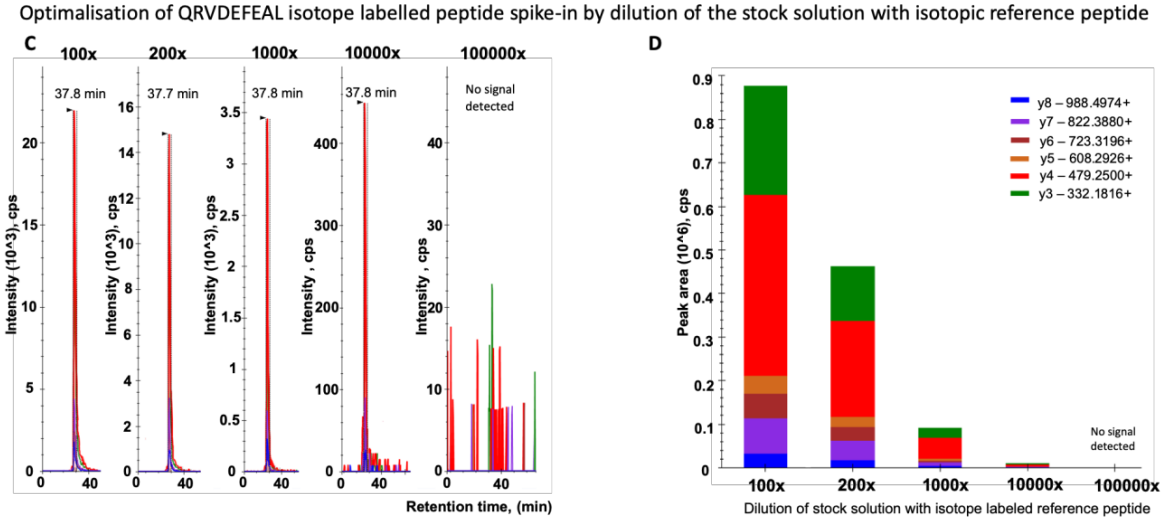


Figure 6. Extracted product ion chromatograms for SRM validated mutated peptides. *(A).* and *(B).* a list of product ions representing two selected high confidence mutated peptides (peptide confidence > 99%) identified using the 2D LC-MS/MS data. Product ions labelled with brown color were identified in the LC-MS/MS spectra of VSGSPEQAVEENLSSYFLDR and IIPTVLMTEDIK peptides (A. and B., respectively). Underlined product ion masses represent identified product ions that confirm presence of the amino acid mutation in the peptide sequence. *(A).* Fragmentation evidence of the VSGSPEQAVEENLSSYFLDR mutated peptide highlights 11 product ions confirming the S to F mutation. *(B).* Fragmentation evidence of the IIPTVLMTEDIK mutated peptide identifies only 3 product ions confirming the A to P mutation. The fragmentation evidences clearly highlight that the probability of amino acid mutation is not reflected in peptide confidence; therefore, these two peptides are shown as an example that it is important to evaluate the spectra/fragmentation evidence to determine mutant sequence status. *(C).* and *(D).* example of pseudo-SRM quantitation of the SIITYVSSLYDTMPR peptide referencing the heavy and light peptide titrations and defining key features required for high confidence validation (1-5) and other peak features (1-3). Left section of *(C).* highlighted in red shows product ions representing light and heavy form of peptide. While product ion chromatograms show MS/MS signal of these product ions across the LC run. *(D).* Tabulated results from a pseudo-SRM validation of selected mutated peptides using isotope labelled reference peptides. The table highlights sequences and charge states of the mutated peptides (the first and the second columns) followed by columns with key characteristics of the product ions representing intrinsic or reference mutated peptides in a sample. The third column defines the consistency of retention times among the reference and intrinsically mutated peptides. The intrinsic peptide must have an identical retention time to the reference peptide as its chromatographic characteristics remain unchanged by isotope labelling. The fourth column defines the intensity of the intrinsic and the reference mutated peptide. Both peptides must have a sufficient intensity (peak height) to be considered as detected. A peptide is defined as detected when its intensity is at least 3 times the background noise of the method (noise of pseudo-SRM was approximately 100 counts per second) according to FDA directions. The fifth column filters out wild-type peptides; the spectra of which could overlay with the spectra of the mutated peptides. We list detected product ions as having the characteristic mass shift of a mutated amino acid; this further validates the existence of a mutated peptide. The higher the number of product ions with the characteristic mass shift of mutated amino acid, the higher the confidence of a mutated peptide. The sixth column describes the similarity of an intensity pattern between the intrinsic and the mutated peptide. The ionization efficiency of both peptides must be the same, therefore we expect the same intensity rank of product ions in both peptides. We inspected corresponding product ion chromatograms and we selected only product ions that show similar patterns in terms of the intensity. The higher the number of product ions the higher the similarity between the intrinsic and the reference mutated peptide. The seventh column shows dotP value. This value reflects and summarizes the similarity between the intrinsic and the reference mutated peptide. It ranges in between 0 to 1 and the higher the value is, the higher the similarity of both peptides. Usually, pseudo-SRM validated peptides have high dotP values. The last column summarizes values determined from the pseudo-SRM assay. Peptides with an excellent pseudo-SRM evidence could be considered as certainly validated. **Abbreviations:** L = (light) naturally occurring intrinsic form of mutated peptide, H = (heavy) reference isotope labelled peptide, WT = wild-type, RT = retention time.

A

TACC3 (S to F) VSGSPEQAVEENLSSYFLDR

Residue	Product Ion series b	b+2	y	y+2
1. V	100.0757	50.5415	2227.0513	1114.0293
2. S	187.1077	94.0575	2127.9829	1064.4951
3. G	244.1292	122.5682	2040.9509	1020.9791
4. S	331.1612	166.0842	1983.9294	992.4684
5. P	428.2140	214.6106	1896.9874	948.9523
6. E	557.2566	279.1319	1799.8446	900.4260
7. Q	685.3151	343.1612	1670.8020	835.9047
8. A	756.3523	378.6798	1542.7435	771.8754
9. V	855.4207	428.2140	1471.7064	736.3568
10. E	984.4633	492.7353	1372.6379	686.8226
11. E	1113.5059	557.2566	1234.5953	622.3013
12. N	1227.5438	614.2780	1114.5527	557.7800
13. L	1340.6329	670.8201	1000.5098	500.7585
14. S	1427.6649	714.3361	887.4258	444.2165
15. S	1514.6969	757.8521	800.3937	400.7005
16. Y	1677.7602	839.3838	713.3617	357.1845
17. F	1824.8286	912.9180	550.2984	275.6528
18. L	1937.9127	969.4600	403.2300	202.1186
19. D	2025.9397	1026.9735	290.1459	145.5766
20. R	2209.0408	1105.0240	175.1190	88.0631

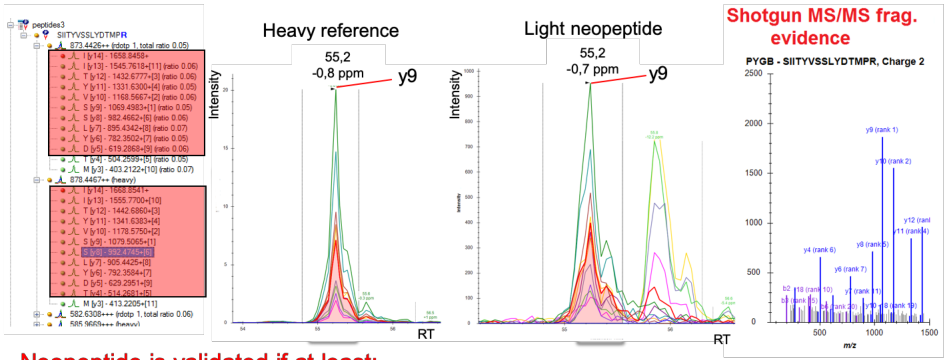
B

EIF4G1 (A to P) IIPTVLMTEDIK

Residue	Product Ion series b	b2+	y	y2+
1. I	114.0913	57.5493	1372.7756	686.8914
2. I	227.1754	1140913	1259.6916	630.3494
3. P	324.2282	162.6177	1146.6075	573.8074
4. T	425.2758	213.1416	1049.5547	525.2810
5. V	524.3443	262.6758	948.5070	474.7572
6. L	637.4283	319.2178	849.4386	425.2230
7. M	768.4688	384.7380	736.3546	368.6809
8. T	869.5165	435.2619	605.3141	303.1607
9. E	998.5591	499.7832	504.2664	252.6368
10. D	1113.5360	557.2967	375.2238	188.1155
11. I	1226.6701	613.8387	260.1969	130.6021
12. K	1354.7551	677.8862	147.1128	74.0600

C

SIITYVSSLYDTMPR (Fraction - 65% B)



Neopeptide is validated if at least:

1. same fraction
2. equal RT
3. Similar fragmentation pattern
4. Similar fragment ion intensities
5. Same m/z shift

and

Other peak-group features:

1. Coelution
2. Shape
3. Summed intensity

D

Mutated peptide candidate identified in 2D LC-MS/MS screen	Precursor ion charge state	RT consistency among H and L mutated peptide (min)	Peak height of the most intensive product ion A375 wt/ A375 p53 null (cps)	Product ions proving the evidence of mutated aminoacid in peptide	Product ions with similar intensity patterns in H and L mutated peptide	dotP value	Overall comment on mutated peptide p-SRM evidence
ASLEAAIADAEQ**H**GELAIK	2+, 3+	yes (46)	1500/1000	7	y12, y11, y7, y6, y10	1	excellent
WEAAH**V**AEQLR	3+	yes (33)	1000/2000	6	y5, y6, y4, y3, y7, y8	1	excellent
YFAGNLASGGAAGATSLCFVYP LDFAR (new trypsin cleavage site before Y)	3+	yes (55)	7000/9000	14	y7, y6	0.89	fair
QRVDEFEA**L**	2+	yes (39)	2500/2500	6	b7, b8, b5, b6, b4, b3	1	excellent
VDEFEA**L**	2+	yes (43)	500/500	4	b6, b5, b3, b4, y3, y4	1	excellent
SIITYVSSLYD**T**MPR	2+, 3+	yes (55)	3000/3000	11	y9, y10, y12, y11	1	excellent
LII**N**LVTSIGDVVNHDPVVGDR	2+, 3+	yes (54)	500/1400	4	y6, y13, y9, y10,	1	excellent
GTAAAAAAAAA**A**K	2+, 3+	yes (32)	7000/3500	2	y10, y9, y11, y8, y7, y6,	1	excellent
VSGSPEQAVEENLSSY**F**LDR	3+	yes (54)	100/200	12	y7, y6	0.89	weak
ILGEVI**N**ALK	2+	yes (44)	2000/2000	6	y9, y8, y6, y5, y4, y7	1	excellent

Figure 7. Generating a melanoma cell line with a p53-null status using CRISP-R mediated gene editing. *(A). Position of guide RNA-binding motif.* Guide RNAs targeting exon 5 of the p53 gene were cloned into pBT-U6-CAS9-2A-GFP. The intron-exon-intron sequence of the p53 gene is highlighted along with the position and orientation of the guide RNA (reverse arrow; in blue with the PAM sequence in red). A375 melanoma cells were transfected with the guide RNA and clones were selected using GFP selection by FACS followed by plating for single cells. Individual cells were grown and screened for evidence of gene editing and p53 activation status. *(B and D). DNA sequencing*. Sequence of the chromosomal DNA derived from the p53 knock-out cells by amplification of the chromosomal region using PCR primers flanking the guide RNA target motif. *(C and E). Translation of the gene edits produces theoretical stop codons*. *(F and G). Immunoblotting to define the p53 and HLA* expression status. (F). Lysates from A375 (wt and p53-null) and HCT116 cells (wt and p53-null) were immunoblotted with the p53 antibody DO-1. This antibody binds to the N-terminal domain of p53. A loading control was detecting PCNA using the PC10 antibody. (G). Cell lysates from the indicated A375 variants (wt and -/-), as SiHa cells (as a control, C) were immunoblotted with antibodies to MHC Class I allele, HLA-B. Actin was used as a loading control.

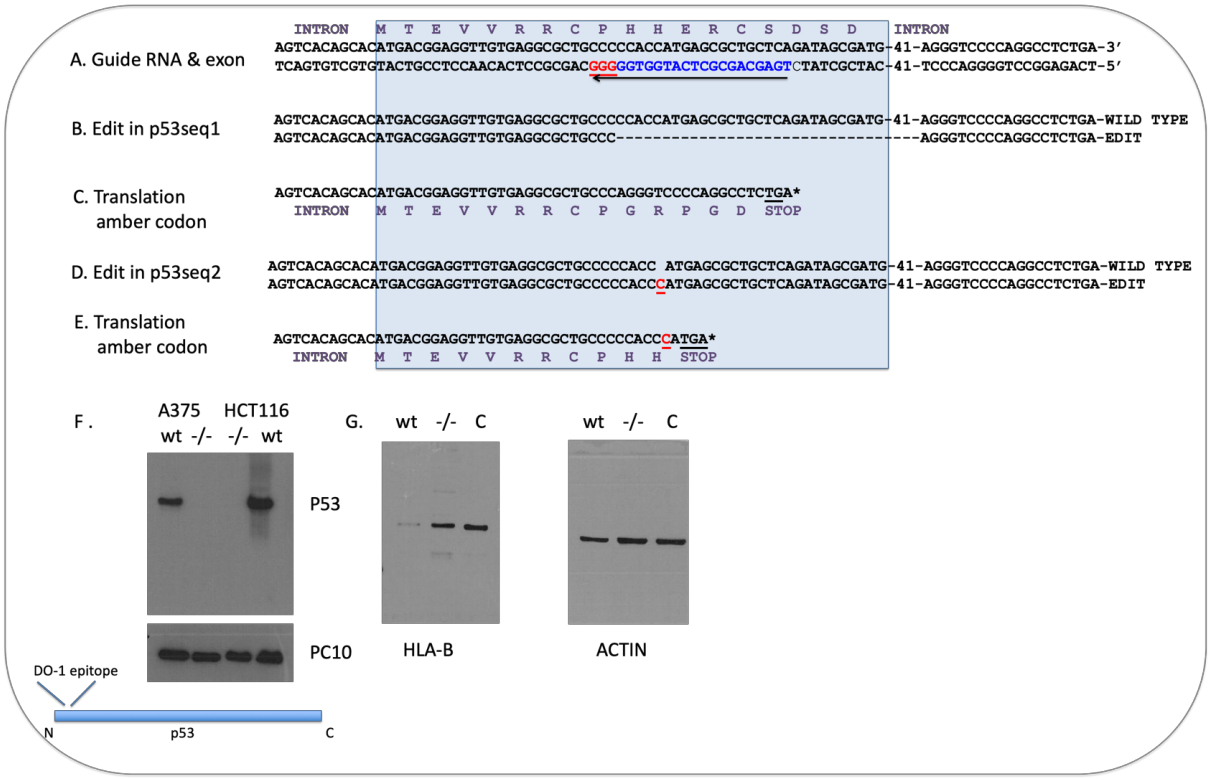


Figure 8. SWATH quantitation pipeline of mutated peptides in isogenic p53 wt and null cell panels. *(A).* An example of extracted product ion chromatogram corresponding to a mutated peptide from Glycogen phosphorylase, brain form (PYGB) detected using SWATH acquisition in A375 cell lysate. A legend in the right shows rank of the most intensive product ions from mutated peptide. *(B).* A spectral library MS/MS spectrum corresponding to peptide LIINLVTSIGDVVNHDPVVGDR. An intensity-based product ion rank in spectral library should be identical to product ion rank in product ion chromatogram to consider peptide hit as valid. *(C).* Quantification using three technical replicates between A375 p53-null cells and A375 p53 wild-type cells shows an up-regulation of mutated peptide in A375 null cells. Here, as problematic we see a quantification on a set of product ions which might refer also to wild-type peptide form. Therefore, we recommend inspecting a set of selected product ions for quantitation and filter out only product ions uniquely covering the mutant position in a peptide sequence. In case of PYGB LIINLVTSIGDVVNHDPVVGDR mutated peptide it is b4 – b21 and y19 – y21 and in case of PLEC mutated peptide SIITYVSSLYDTMPR it is b12 – b15 and y4 – y15. *(D).* Extraction of product ion chromatogram for SIITYVSSLYDTMPR mutated peptide (PLEC). SIITYVSSLYDTMPR quantitation refers more uniquely to the mutated peptide form compared to quantitation of LIINLVTSIGDVVNHDPVVGDR peptide. It relies on product ions that are characteristic exclusively for mutant peptide form (y4, y5, y6, y7). *(E).* Shows corresponding spectral library for SIITYVSSLYDTMPR mutated peptide and intensity rank of product ions. *(F).* Shows quantitation of SIITYVSSLYDTMPR mutated peptide in three technical replicates of A375 p53-null cells and A375 p53 wild-type cells. Relative quantitation of both mutated peptides in A375 p53- and A375 WT cell line *(C).* and *(F).* was determined from a sum of 4 product ion peak areas by integration of AUC. Bar graph *(C* and *F).* represents peak areas corresponding to these 4 product ions. Overall peptide quantity in each replicate is represented by an entire bar. *(G).* SWATH quantitation of a control peptide from ACTB to evaluate the effects of sample preparation and differences between mass spectrometry measurements.

Mutated peptide quantitation using SWATH acquisition in A375 cell lines

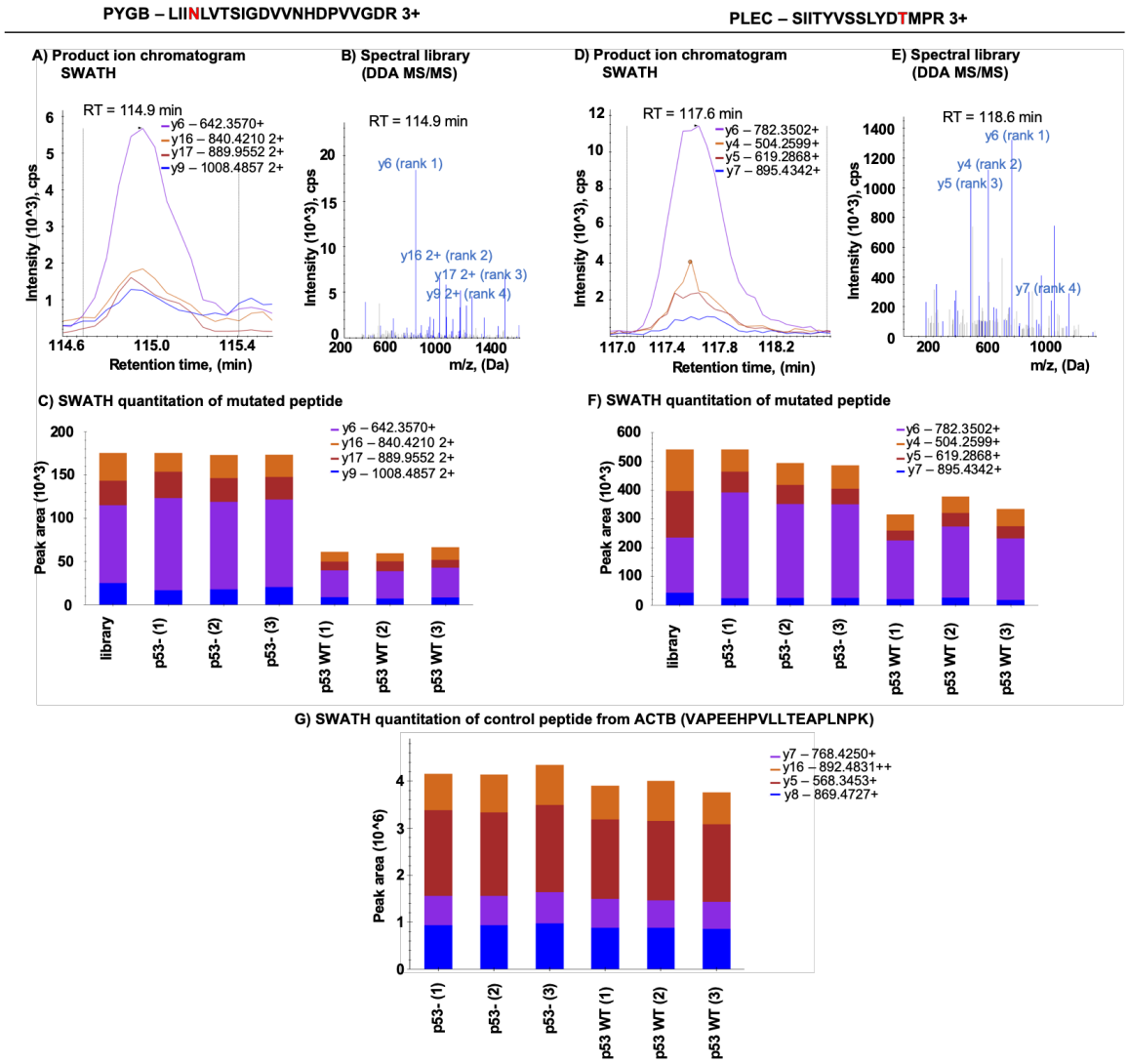


Figure 9. Summary of mutated peptides detected using the optimized platforms. Using both CLCbio (CLC) and Varscan (VS), variants were detected requiring at least 10 RNA mutant reads and 1 DNA mutant read. A. A comparison of the enriched mutant peptides detected in wt-p53 vs p53-null cells using mass spectral data summarized in Supplementary Tables 13 and 14. B and C. Summarizes the mutant tryptic peptides detected in wt-p53 or p53-null cells, using CLCbio and Varscan platforms. The reference database generated from the DNA and RNAseq was converted to protein amino acid sequences by TransPEM [47] and subsequently this reference database was used in ProteinPilot 4.5 to search for spectra with matches that define mutated peptides.

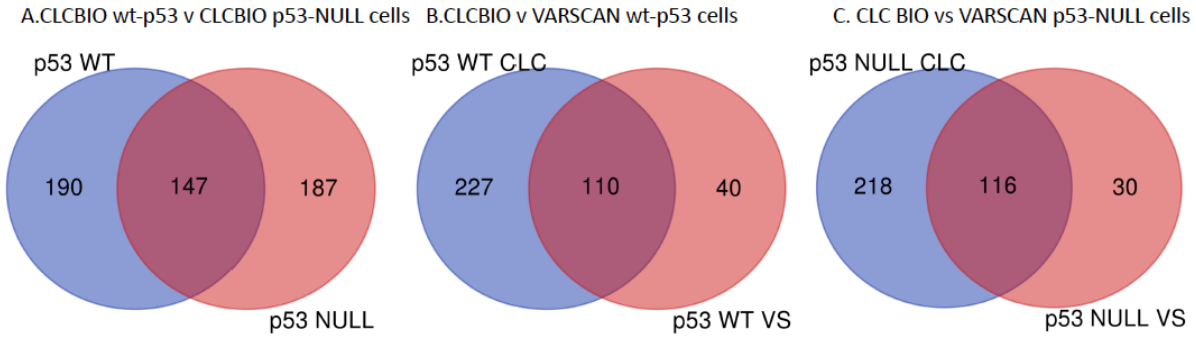


Figure 10. STRING analysis of mutant protein networks in wt-p53 cells. STRING (https://string-db.org) is a database of protein-protein interactions including direct experimental data, indirect functional associations, from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from additional primary databases [48]. The known interactions are defined as: curated databases (——), and experimentally determined (——). The predicted interactions are defined as: gene neighborhoods (——), gene fusions (——), gene co-occurrence (——), textmining (——), co-expression (——), and protein homology (——). The minimum required interaction score was defined as high confidence (0.7). STRING was used to define interaction networks composed of mutant proteins in the wt-p53 cell tryptic peptide dataset that differ from p53-null cells. The mutant peptides from wt-p53 cells are summarized in Supplementary Table 13.

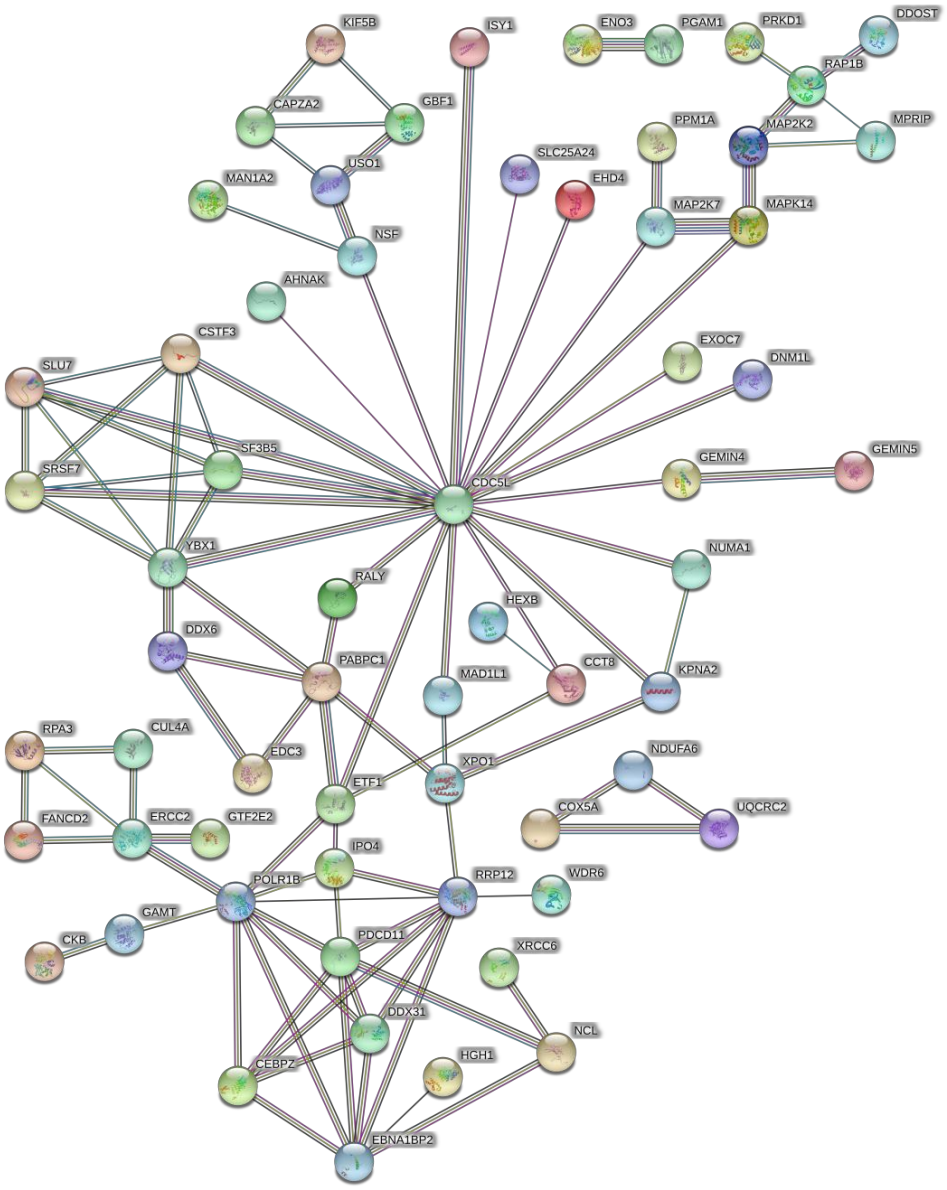


Figure 11. STRING analysis of mutant protein networks in p53-null cells. STRING was used as in Fig. 10 to define interaction networks composed of mutant proteins in the p53-null cell tryptic peptide dataset that differs from wt-p53 cells. The known interactions are defined as: curated databases (——), and experimentally determined (——). The predicted interactions are defined as: gene neighborhoods (——), gene fusions (——), gene co-occurance (——), textmining (——), co-expression (——), and protein homology (——). The minimum required interaction score was defined as high confidence (0.7). The mutant peptides from p53-null cells are summarized in Supplementary Table 14.

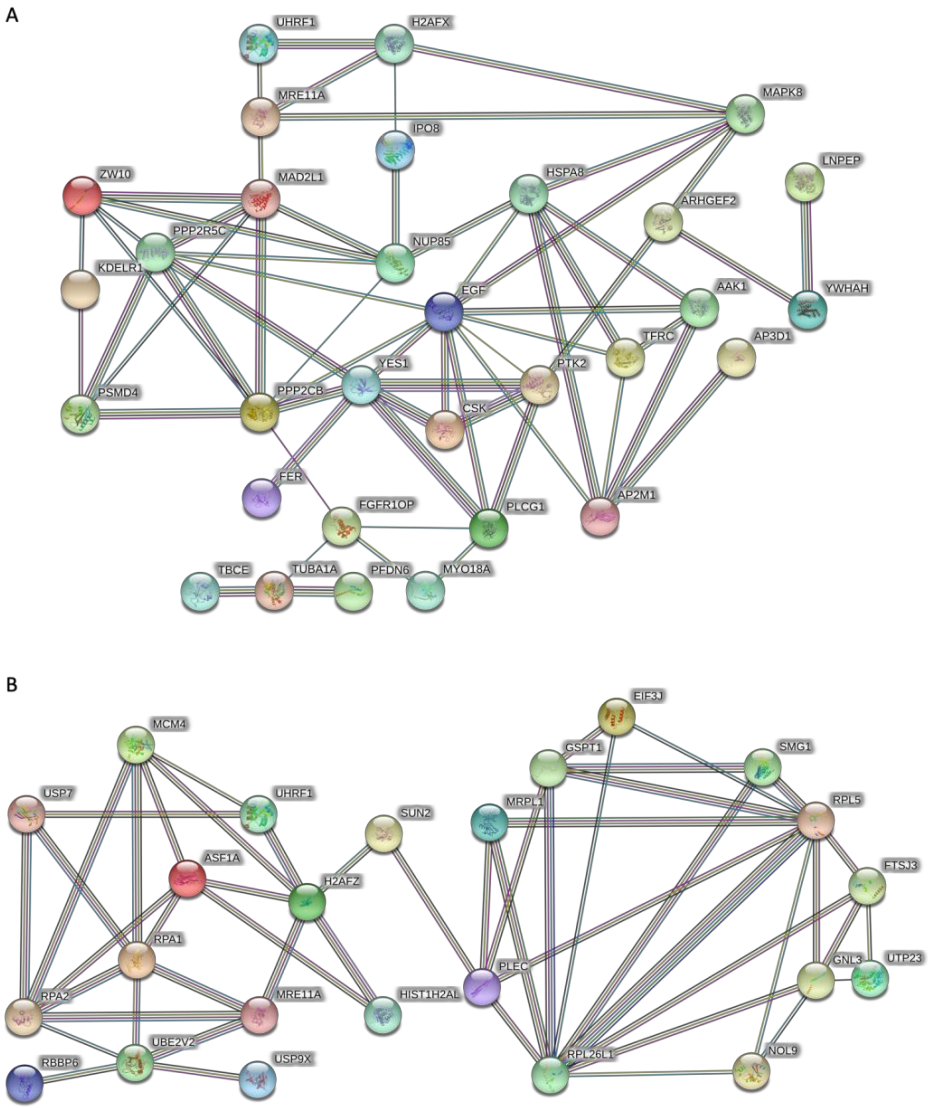


TABLE 1. SWATH-MS analysis of proteins whose genes are mutated using wt-p53 and p53-null A375 cells. The table highlights gene name, amino acid substitution, and fold change in the expression of wt vs p53-null cells. Only one protein on this list has a mutated peptide detected using mass spectrometry (rpl14, Fig. 3). The other 18 proteins with homozygous mutations do not have any tryptic mutated peptides detected using SWATH-MS.

N	Gene name	Amino acid change(s)	Fold Change	N	Gene name	Amino acid change(s)	Fold Change
1	**SEC31A**	Asp22Tyr	**2.74**	22	**PRKCSH**	Pro33Leu	**0.98**
2	**PYGB**	Lys622Asn	**2.42**	23	**TCP1**	Lys321Arg	**0.98**
3	**ADPGK**	Ala8Val	**2.34**	24	**SPTBN1**	Pro883Thr	**0.98**
4	**EGFR**	Arg521Lys	**2.29**	25	**PABPC1**	Leu593Val Leu597Pro Met584Ile Leu562Ser Val517Leu Arg506Cys Val505Ile Arg493Cys Pro402Leu Arg374Cys Glu372Gly Met251Ile Ala154Gly Thr147Met His144Arg Cys132Gly	**0.97**
5	**LYAR**	His265Arg (275 in this transcript)	**2.11**	26	**PSMA2**	Glu3Lys	**0.97**
6	**NDUFA10**	Phe169Leu	**1.59**	27	**PDHA1**	Arg27fs	**0.96**
7	**HADH**	Leu90Pro	**1.52**	28	**HSP90AB1**	Arg719His	**0.93**
8	**LAMC2**	Asp784Asn	**1.33**	29	**MARS**	Thr32Ser	**0.93**
9	**LRRC59**	Glu69Asp Lys137_Pro138insGln	**1.29**	30	**PPP1CA**	Phe225fs	**0.92**
10	**SLC25A5**	Gly73Ser Asn77Thr	**1.23**	31	**NCL**	Asp258del	**0.91**
11	**BCLAF1**	Thr888Asn	**1.19**	32	**ACTG1**	Ser185Cys Glu170Val	**0,90**
12	**WDR1**	Ile185Val	**1.16**	33	**PRSS1**	Asn29Ile Cys185Tyr Ser195Asn Met197Val	**0.89**
13	**SNRPC**	Cys13_Arg14insHis Arg14Leu	**1.14**	34	**RPL14**	Ala159_Lys160insAlaAla	**0.86**
14	**PDCD6IP**	Val378Ile	**1.13**	35	**AHCY**	Arg257Gln	**0.85**
15	**NIT2**	Phe27Ser	**1.10**	36	**SRRM2**	Arg581Trp	**0.82**
16	**PHB2**	Arg123Cys	**1.09**	37	**FUBP1**	Gly455Trp	**0.80**
17	**HEATR1**	**Val1854Ala**	**1.09**	38	**PPA2**	Lys282Asn	**0.72**
18	**RBMX**	**Asn1Lys**	**1.05**	39	**ARHGEF2**	Gly582Trp	**0.64**
19	**NACA**	**Pro649Ser**	**1.03**	40	**FAM162A**	Arg142Cys	**0.54**
20	**PFDN5**	**Pro65Ser**	**1.03**	41	**GC**	His445Arg	**0.40**
21	**PLEC**	**Ala398Thr**	**0,99**	42	**MPRIP**	Ser189del	**0.20**

SUPPLEMENTAL LEGENDS

Supplementary Table 1A. Excel file showing a list of mutated genes with mutations detected at a 5% frequency or higher in DNA from A375 cells using the *CLCbio* variant platform detector. The analysis identified 120,325 genes. Represented are data including chromosome number and position, reference allele and mutation type, zygosity, Count (mutation number) and coverage (total sequencing reads), frequency of mutation, gene cards name, ensemble qualifiers, amino acid change, coding region change (Syn or non-syn), and dbSNP identifiers. **Supplementary Table 1B**. Excel file showing a list of mutated genes detected at a 40% frequency or higher in DNA from A375 cells (from Supplementary Table 1A) using the *CLCbio* variant detector. This analysis generated 63,880 variants with mutations. The data are plotted as a function of; chromosome; region; type of mutation; reference base; allele base; reference allele; zygosity; count; coverage; frequency of mutation; gene card name; ensemble name; gene name; gene biotype; transcript name; coding region change; amino acid change; amino acid change in longest transcript; other variants in codon; whether the mutation is non-synonymous, synonymous, or out with exons; and dbsnp reference.

Supplementary Table 2. Excel file showing a list of mutated genes with non-synonymous mutations detected at a 40% frequency or higher in A375 cells (from Supplementary Table 1B) using the *CLCbio* variant detector. This analysis generated 1,468 genes with non-synonymous mutations. The data are plotted as a function of; chromosome; region; type of mutation; reference base; allele base; reference allele; length of change; zygosity; count; coverage; frequency of mutation; probability; forward read count; reverse read count; ratio; average quality; exact match to dbsnp; gene card name; ensembl name; gene name; gene biotype; transcript name; coding region change; amino acid change; amino acid change in longest transcript; other variants in codon; whether the mutation is non-synonymous, synonymous, or out with exons; and dbsnp reference.

Supplementary Table 3. Excel file showing RNA sequencing reads mapped to the human reference genome hg19, with dbSNPs removed, to identify mutated and expressed genes. A total of 18,341 expressed non-synonymous mRNA variants were identified with a cutoff of 5% of the total RNA reads defined as mutated. The data are plotted as a function of; chromosome; position; type of mutation; reference; allele; reference allele; length of change; zygosity; RNA count; RNA coverage; RNA frequency; RNA forward reads; RNA reverse reads; DNA read count; DNA read coverage; gene card name; ensemble name; gene name; transcript ID; coding region change; amino acid change; and whether the mutation is non-synonymous, synonymous, or out with exons.

Supplementary Table 4. Excel file showing RNAseq data with a stringent cutoff requiring at least 40 mutated RNA sequencing reads and filtered against genomic DNA cutoff of at least 1 mutant genomic DNA sequencing read. This produced 5,980 RNA variants including synonymous, non-synonymous, and non-exonic mutations. The data are plotted as a function of; chromosome number; type of mutation; zygosity; gene card name; ensemble name; gene name; amino acid change; and whether the mutation is non-synonymous, synonymous, or out with exons.

Supplementary Table 5. Excel file showing RNAseq data with a stringent cutoff requiring at least 40 mutated RNA sequencing reads and filtered against genomic DNA cutoff of at least 1 mutant genomic DNA sequencing read (from Supplementary Table 4). Upon filtering for non-synonymous variants, a list was generated composed of 1,418 non-synonymous highly expressed RNA variants. This provides a conservative estimate of the number of mutated expressed mRNA in the A375 cell

line, as there are many mutated mRNAs quantified at reads from 39 and lower. However, we focus on those mutant mRNA species which are abundant, perhaps not degraded by NMD or perhaps not resulting from expression of minor subclones in the A375 cell population.

Supplementary Table 6. Excel file showing a subset of genes acquired after the list composed of 1,418 non-synonymous highly expressed RNA variants (Supplementary Table 5) with the 1,468 CLC genomic DNA variants using the stringent DNA variant calling (Supplementary Table 2). The excel file highlights a total of 877 out of 985 genes with expressed mRNA variants are not present in the original DNA variant list.

Supplementary Table 7. Excel file showing the tryptic peptides processed using SWATH mass spectrometry to identify proteins differentially expressed in the wt and p53-null cells using the *normal* reference proteome Swiss-prot and TrEMBL (potential mutant proteins from this are listed in Table 1).

Supplementary Table 8. Excel file showing tryptic peptides identified from shotgun mass spectrometry including a pre-fractionation step to increase the coverage of total peptides to over 35,000 with a coverage of over 4,500 proteins. (A). represents data from wt-A375 cells and (B). represents data from p53-null A375 cells.

Supplementary Table 9. Excel file showing a subset of proteins listed by selecting proteins from the expressed (mRNA) and mutated genes detectable in the shotgun 2D LC-MS/MS experiment (Supplementary Table 8A and 8B) using the RNA variant file (Supplementary Table 5). Applying RNAseq derived mutant search database to search the 2D LC-MS/MS data, this increased the number of detectable mutant peptides to 193. Of this selection, we manually validated all spectra from the 193 peptides and we produced data (Fig. 4) of 60 mutated proteins of relatively high confidence. Of these 60 mutant proteins, 10 were validated by SRM (Fig. 5; Supplementary Table 10) to produce a conservative FDR of 20%. Thus, 4 out of 5 of the manually validated, high confidence mutated peptides are likely to be mutated.

Supplementary Table 10. Excel file showing results summary of SRM mass spectrometry analysis in a subset of 10 mutated proteins. Evaluating similarity of product ion peakgroups referring to neopeptides is summarised by dotp value ranging from 0 to 1. Neopeptides with dotp above 0.9 and with more than 2 product ions above LOD (three times noise) are considered as successfully validated. Abbreviation H refers to spiked-in "heavy" isotope labelled neopeptide and L refers to endogenous "Light" form of neopeptide.

Supplementary Table 11. A list of DNA variants detected using VarScan2 using a minimal coverage of 1 and hg38.
Supplementary Table 12. A list of RNA variants detected using VarScan2 using a minimal coverage of 10 and hg38.
Supplementary Table 13. Mutant tryptic peptide lists derived from lysates from wt-p53 cells using CLCbio software.
Supplementary Table 14. Mutant tryptic peptide lists derived from lysates from p53-null cells using CLCbio software.

REFERENCES

[1] G.R. Bignell, C.D. Greenman, H. Davies, A.P. Butler, S. Edkins, J.M. Andrews, G. Buck, L. Chen, D. Beare, C. Latimer, S. Widaa, J. Hinton, C. Fahey, B. Fu, S. Swamy, G.L. Dalgliesh, B.T. Teh, P. Deloukas, F. Yang, P.J. Campbell, P.A. Futreal, M.R. Stratton, Signatures of mutation and selection in the cancer genome, Nature, 463 (2010) 893-898.
[2] M.R. Stratton, Journeys into the genome of cancer cells, EMBO Mol Med, 5 (2013) 169-172.
[3] J.W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical R&D efficiency, Nat Rev Drug Discov, 11 (2012) 191-200.
[4] M. Paoloni, C. Khanna, Translation of new cancer treatments from pet dogs to humans, Nat Rev Cancer, 8 (2008) 147-156.
[5] L.B. Alexandrov, P.H. Jones, D.C. Wedge, J.E. Sale, P.J. Campbell, S. Nik-Zainal, M.R. Stratton, Clock-like mutational processes in human somatic cells, Nat Genet, 47 (2015) 1402-1407.
[6] S. Nik-Zainal, J.E. Kucab, S. Morganella, D. Glodzik, L.B. Alexandrov, V.M. Arlt, A. Weninger, M. Hollstein, M.R. Stratton, D.H. Phillips, The genome as a record of environmental exposure, Mutagenesis, 30 (2015) 763-770.
[7] B. An, T. Pan, J. Hu, Y. Pang, L. Huang, A.S.C. Chan, X. Li, J. Yan, The discovery of a potent and selective third-generation EGFR kinase inhibitor as a therapy for EGFR L858R/T790M double mutant non-small cell lung cancer, Eur J Med Chem, 183 (2019) 111709.
[8] J.P. Allison, Checkpoints, Cell, 162 (2015) 1202-1205.
[9] T. Schumacher, L. Bunse, S. Pusch, F. Sahm, B. Wiestler, J. Quandt, O. Menn, M. Osswald, I. Oezen, M. Ott, M. Keil, J. Balss, K. Rauschenbach, A.K. Grabowska, I. Vogler, J. Diekmann, N. Trautwein, S.B. Eichmuller, J. Okun, S. Stevanovic, A.B. Riemer, U. Sahin, M.A. Friese, P. Beckhove, A. von Deimling, W. Wick, M. Platten, A vaccine targeting mutant IDH1 induces antitumour immunity, Nature, 512 (2014) 324-327.
[10] B.M. Carreno, V. Magrini, M. Becker-Hapak, S. Kaabinejadian, J. Hundal, A.A. Petti, A. Ly, W.R. Lie, W.H. Hildebrand, E.R. Mardis, G.P. Linette, Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells, Science, 348 (2015) 803-808.
[11] T. Schlake, A. Thess, M. Fotin-Mleczek, K.J. Kallen, Developing mRNA-vaccine technologies, RNA Biol, 9 (2012) 1319-1330.
[12] S. Colloca, E. Barnes, A. Folgori, V. Ammendola, S. Capone, A. Cirillo, L. Siani, M. Naddeo, F. Grazioli, M.L. Esposito, M. Ambrosio, A. Sparacino, M. Bartiromo, A. Meola, K. Smith, A. Kurioka, G.A. O'Hara, K.J. Ewer, N. Anagnostou, C. Bliss, A.V. Hill, C. Traboni, P. Klenerman, R. Cortese, A. Nicosia, Vaccine vectors derived from a large collection of simian adenoviruses induce potent cellular immunity across multiple species, Sci Transl Med, 4 (2012) 115ra112.
[13] B. Zhang, J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M.C. Chambers, L.J. Zimmerman, K.F. Shaddox, S. Kim, S.R. Davies, S. Wang, P. Wang, C.R. Kinsinger, R.C. Rivers, H. Rodriguez, R.R. Townsend, M.J. Ellis, S.A. Carr, D.L. Tabb, R.J. Coffey, R.J. Slebos, D.C. Liebler, C. Nci, Proteogenomic characterization of human colon and rectal cancer, Nature, 513 (2014) 382-387.
[14] K.V. Ruggles, K. Krug, X. Wang, K.R. Clauser, J. Wang, S.H. Payne, D. Fenyo, B. Zhang, D.R. Mani, Methods, Tools and Current Perspectives in Proteogenomics, Mol Cell Proteomics, 16 (2017) 959-981.
[15] S. Apcher, C. Daskalogianni, R. Fahraeus, Pioneer translation products as an alternative source for MHC-I antigenic peptides, Mol Immunol, 68 (2015) 68-71.
[16] J. Crappe, E. Ndah, A. Koch, S. Steyaert, D. Gawron, S. De Keulenaer, E. De Meester, T. De Meyer, W. Van Criekinge, P. Van Damme, G. Menschaert, PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration, Nucleic Acids Res, 43 (2015) e29.
[17] G.S. Krasnov, A.A. Dmitriev, A.V. Kudryavtseva, A.V. Shargunov, D.S. Karpov, L.A. Uroshlev, N.V. Melnikova, V.M. Blinov, E.V. Poverennaya, A.I. Archakov, A.V. Lisitsa, E.A. Ponomarenko,

PPLine: An Automated Pipeline for SNP, SAP, and Splice Variant Detection in the Context of Proteogenomics, J Proteome Res, 14 (2015) 3729-3737.
[18] P. Jagtap, J. Goslinga, J.A. Kooren, T. McGowan, M.S. Wroblewski, S.L. Seymour, T.J. Griffin, A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies, Proteomics, 13 (2013) 1352-1357.
[19] B. Afsari, T. Guo, M. Considine, L. Florea, L.T. Kagohara, G.L. Stein-O'Brien, D. Kelley, E. Flam, K.D. Zambo, P.K. Ha, D. Geman, M.F. Ochs, J.A. Califano, D.A. Gaykalova, A.V. Favorov, E.J. Fertig, Splice Expression Variation Analysis (SEVA) for inter-tumor heterogeneity of gene isoform usage in cancer, Bioinformatics, 34 (2018) 1859-1867.
[20] B. Wen, S. Xu, G.M. Sheynkman, Q. Feng, L. Lin, Q. Wang, X. Xu, J. Wang, S. Liu, sapFinder: an R/Bioconductor package for detection of variant peptides in shotgun proteomics experiments, Bioinformatics, 30 (2014) 3136-3138.
[21] F. Zickmann, B.Y. Renard, MSProGene: integrative proteogenomics beyond six-frames and single nucleotide polymorphisms, Bioinformatics, 31 (2015) i106-115.
[22] J. Hedegaard, K. Thorsen, M.K. Lund, A.M. Hein, S.J. Hamilton-Dutoit, S. Vang, I. Nordentoft, K. Birkenkamp-Demtroder, M. Kruhoffer, H. Hager, B. Knudsen, C.L. Andersen, K.D. Sorensen, J.S. Pedersen, T.F. Orntoft, L. Dyrskjot, Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue, PLoS One, 9 (2014) e98187.
[23] V. Tiedje, S. Ting, T. Herold, S. Synoracki, S. Latteyer, L.C. Moeller, D. Zwanziger, M. Stuschke, D. Fuehrer, K.W. Schmid, NGS based identification of mutational hotspots for targeted therapy in anaplastic thyroid carcinoma, Oncotarget, 8 (2017) 42613-42620.
[24] V. Gorshkov, T. Verano-Braga, F. Kjeldsen, SuperQuant: A Data Processing Approach to Increase Quantitative Proteome Coverage, Anal Chem, 87 (2015) 6319-6327.
[25] I.F. do Valle, E. Giampieri, G. Simonetti, A. Padella, M. Manfrini, A. Ferrari, C. Papayannidis, I. Zironi, M. Garonzi, S. Bernardi, M. Delledonne, G. Martinelli, D. Remondini, G. Castellani, Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data, BMC Bioinformatics, 17 (2016) 341.
[26] J.P. Blaydes, M.G. Luciani, S. Pospisilova, H.M. Ball, B. Vojtesek, T.R. Hupp, Stoichiometric phosphorylation of human p53 at Ser315 stimulates p53-dependent transcription, J Biol Chem, 276 (2001) 4699-4708.
[27] J.P. Blaydes, T.R. Hupp, DNA damage triggers DRB-resistant phosphorylation of human p53 at the CK2 site, Oncogene, 17 (1998) 1045-1052.
[28] S.P. Shah, A. Roth, R. Goya, A. Oloumi, G. Ha, Y. Zhao, G. Turashvili, J. Ding, K. Tse, G. Haffari, A. Bashashati, L.M. Prentice, J. Khattra, A. Burleigh, D. Yap, V. Bernard, A. McPherson, K. Shumansky, A. Crisan, R. Giuliany, A. Heravi-Moussavi, J. Rosner, D. Lai, I. Birol, R. Varhol, A. Tam, N. Dhalla, T. Zeng, K. Ma, S.K. Chan, M. Griffith, A. Moradian, S.W. Cheng, G.B. Morin, P. Watson, K. Gelmon, S. Chia, S.F. Chin, C. Curtis, O.M. Rueda, P.D. Pharoah, S. Damaraju, J. Mackey, K. Hoon, T. Harkins, V. Tadigotla, M. Sigaroudinia, P. Gascard, T. Tlsty, J.F. Costello, I.M. Meyer, C.J. Eaves, W.W. Wasserman, S. Jones, D. Huntsman, M. Hirst, C. Caldas, M.A. Marra, S. Aparicio, The clonal and mutational evolution spectrum of primary triple-negative breast cancers, Nature, 486 (2012) 395-399.
[29] R.D. Morin, M. Mendez-Lago, A.J. Mungall, R. Goya, K.L. Mungall, R.D. Corbett, N.A. Johnson, T.M. Severson, R. Chiu, M. Field, S. Jackman, M. Krzywinski, D.W. Scott, D.L. Trinh, J. Tamura-Wells, S. Li, M.R. Firme, S. Rogic, M. Griffith, S. Chan, O. Yakovenko, I.M. Meyer, E.Y. Zhao, D. Smailus, M. Moksa, S. Chittaranjan, L. Rimsza, A. Brooks-Wilson, J.J. Spinelli, S. Ben-Neriah, B. Meissner, B. Woolcock, M. Boyle, H. McDonald, A. Tam, Y. Zhao, A. Delaney, T. Zeng, K. Tse, Y. Butterfield, I. Birol, R. Holt, J. Schein, D.E. Horsman, R. Moore, S.J. Jones, J.M. Connors, M. Hirst, R.D. Gascoyne, M.A. Marra, Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma, Nature, 476 (2011) 298-303.

[30] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, Nat Methods, 6 (2009) 359-362.
[31] B.C. Collins, C.L. Hunter, Y. Liu, B. Schilling, G. Rosenberger, S.L. Bader, D.W. Chan, B.W. Gibson, A.C. Gingras, J.M. Held, M. Hirayama-Kurogi, G. Hou, C. Krisp, B. Larsen, L. Lin, S. Liu, M.P. Molloy, R.L. Moritz, S. Ohtsuki, R. Schlapbach, N. Selevsek, S.N. Thomas, S.C. Tzeng, H. Zhang, R. Aebersold, Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry, Nat Commun, 8 (2017) 291.
[32] L. Ding, M. Kim, K.L. Kanchi, N.D. Dees, C. Lu, M. Griffith, D. Fenstermacher, H. Sung, C.A. Miller, B. Goetz, M.C. Wendl, O. Griffith, L.A. Cornelius, G.P. Linette, J.F. McMichael, V.K. Sondak, R.C. Fields, T.J. Ley, J.J. Mule, R.K. Wilson, J.S. Weber, Clonal architectures and driver mutations in metastatic melanomas, PLoS One, 9 (2014) e111153.
[33] P.A. Ott, Z. Hu, D.B. Keskin, S.A. Shukla, J. Sun, D.J. Bozym, W. Zhang, A. Luoma, A. Giobbie-Hurder, L. Peter, C. Chen, O. Olive, T.A. Carter, S. Li, D.J. Lieb, T. Eisenhaure, E. Gjini, J. Stevens, W.J. Lane, I. Javeri, K. Nellaiappan, A.M. Salazar, H. Daley, M. Seaman, E.I. Buchbinder, C.H. Yoon, M. Harden, N. Lennon, S. Gabriel, S.J. Rodig, D.H. Barouch, J.C. Aster, G. Getz, K. Wucherpfennig, D. Neuberg, J. Ritz, E.S. Lander, E.F. Fritsch, N. Hacohen, C.J. Wu, An immunogenic personal neoantigen vaccine for patients with melanoma, Nature, 547 (2017) 217-221.
[34] D.P. Lane, Cancer. p53, guardian of the genome, Nature, 358 (1992) 15-16.
[35] K. Edlund, O. Larsson, A. Ameur, I. Bunikis, U. Gyllensten, B. Leroy, M. Sundstrom, P. Micke, J. Botling, T. Soussi, Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultradeep sequencing of human tumors, Proc Natl Acad Sci U S A, 109 (2012) 9551-9556.
[36] M. Overholtzer, P.H. Rao, R. Favis, X.Y. Lu, M.B. Elowitz, F. Barany, M. Ladanyi, R. Gorlick, A.J. Levine, The presence of p53 mutations in human osteosarcomas correlates with high levels of genomic instability, Proc Natl Acad Sci U S A, 100 (2003) 11547-11552.
[37] M.S. Lawrence, P. Stojanov, P. Polak, G.V. Kryukov, K. Cibulskis, A. Sivachenko, S.L. Carter, C. Stewart, C.H. Mermel, S.A. Roberts, A. Kiezun, P.S. Hammerman, A. McKenna, Y. Drier, L. Zou, A.H. Ramos, T.J. Pugh, N. Stransky, E. Helman, J. Kim, C. Sougnez, L. Ambrogio, E. Nickerson, E. Shefler, M.L. Cortes, D. Auclair, G. Saksena, D. Voet, M. Noble, D. DiCara, P. Lin, L. Lichtenstein, D.I. Heiman, T. Fennell, M. Imielinski, B. Hernandez, E. Hodis, S. Baca, A.M. Dulak, J. Lohr, D.A. Landau, C.J. Wu, J. Melendez-Zajgla, A. Hidalgo-Miranda, A. Koren, S.A. McCarroll, J. Mora, B. Crompton, R. Onofrio, M. Parkin, W. Winckler, K. Ardlie, S.B. Gabriel, C.W.M. Roberts, J.A. Biegel, K. Stegmaier, A.J. Bass, L.A. Garraway, M. Meyerson, T.R. Golub, D.A. Gordenin, S. Sunyaev, E.S. Lander, G. Getz, Mutational heterogeneity in cancer and the search for new cancer-associated genes, Nature, 499 (2013) 214-218.
[38] J.M.J. Weaver, C.S. Ross-Innes, N. Shannon, A.G. Lynch, T. Forshew, M. Barbera, M. Murtaza, C.J. Ong, P. Lao-Sirieix, M.J. Dunning, L. Smith, M.L. Smith, C.L. Anderson, B. Carvalho, M. O'Donovan, T.J. Underwood, A.P. May, N. Grehan, R. Hardwick, J. Davies, A. Oloumi, S. Aparicio, C. Caldas, M.D. Eldridge, P.A.W. Edwards, N. Rosenfeld, S. Tavare, R.C. Fitzgerald, O. consortium, Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis, Nat Genet, 46 (2014) 837-843.
[39] E. Garcia, A. Hayden, C. Birts, E. Britton, A. Cowie, K. Pickard, M. Mellone, C. Choh, M. Derouet, P. Duriez, F. Noble, M.J. White, J.N. Primrose, J.C. Strefford, M. Rose-Zerilli, G.J. Thomas, Y. Ang, A.D. Sharrocks, R.C. Fitzgerald, T.J. Underwood, O. consortium, Authentication and characterisation of a new oesophageal adenocarcinoma cell line: MFD-1, Sci Rep, 6 (2016) 32417.
[40] M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system, Bioinformatics, 32 (2016) 511-517.
[41] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system, Nat Protoc, 8 (2013) 2281-2308.

[42] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol, 14 (2013) R36.
[43] D.C. Koboldt, Q. Zhang, D.E. Larson, D. Shen, M.D. McLellan, L. Lin, C.A. Miller, E.R. Mardis, L. Ding, R.K. Wilson, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing, Genome Res, 22 (2012) 568-576.
[44] T.S. Batth, C. Francavilla, J.V. Olsen, Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics, J Proteome Res, 13 (2014) 6176-6186.
[45] L.C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, Mol Cell Proteomics, 11 (2012) O111 016717.
[46] L. Way, J. Faktor, P. Dvorakova, J. Nicholson, B. Vojtesek, D. Graham, K.L. Ball, T. Hupp, Rearrangement of mitochondrial pyruvate dehydrogenase subunit dihydrolipoamide dehydrogenase protein-protein interactions by the MDM2 ligand nutlin-3, Proteomics, 16 (2016) 2327-2344.
[47] F. Zavadil Kokas, J. Faktor, B. Vojtesek, Cooperation of Genomic, Transcriptomics and Proteomic Methods in the Detection of Mutated Proteins, Klin Onkol, 32 (2019) 78-84.
[48] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res, 47 (2019) D607-D613.

The effects of p53 gene inactivation on mutant proteome expression in a human melanoma cell model

Jakub Faktor[2], Giuseppa Grasso[1+], Filip Zavadil[2], Małgorzata Kurkowiak[3], Marcos Yébenes Mayordomo[1,3], Sachin Kote[3], Ashita Singh[1], Li Ruidong[1,5], J. Robert O'Neill[1, 6], Petr Muller[2], David Goodlett[3,4], Borek Vojtesek[2], and Ted Hupp[1,2,3]

[1]University of Edinburgh, Institute of Genetics and Molecular Medicine, Edinburgh, UK; [2]Research Centre for Applied Molecular Oncology (RECAMO), Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic; [3]University of Gdansk, International Centre for Cancer Vaccine Science, ul. Wita Stwosza 63, 80-308 Gdansk, Poland; [4]School of Pharmacy, University of Maryland, Baltimore, USA ICCVS; [5]Department of Gastrointestinal Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China. [6]University of Cambridge, Cambridge, UK. + current address, Institut de Génétique Humaine, CNRS, University of Montpellier, Gene Regulation Laboratory, 141 rue de la cardonille, Montpellier, France.

Running title: An isogenic p53-null melanoma cell model for use in mutant proteomics

Keywords: cancer, p53, protein mass spectrometry, proteogenomics, proteomics

Correspondence, vojtesek@mou.cz , ted.hupp@ed.ac.uk

ABSTRACT

Background: The identification of mutated proteins in human cancer cells-termed proteogenomics, requires several technologically independent research methodologies including DNA variant identification, RNA sequencing, and mass spectrometry. Any one of these methodologies are not optimized for identifying potential mutated proteins and any one output fails to cover completely a specific landscape.

Methods: An isogenic melanoma cell with a p53-null genotype was created by CRISPR/CAS9 system to determine how p53 gene inactivation affects mutant proteome expression. A mutant peptide reference database was developed by comparing two distinct DNA and RNA variant detection platforms using these isogenic cells. Chemically fractionated tryptic peptides from lysates were processed using a TripleTOF 5600+ mass spectrometer and their spectra were identified against this mutant reference database.

Results: Approximately 190 mutated peptides were enriched in wt-p53 cells, 187 mutant peptides were enriched in p53-null cells, with an overlap of 147 mutated peptides. STRING analysis highlighted that the wt-p53 cell line was enriched for mutant protein pathways such as CDC5L and POLR1B, whilst the p53-null cell line was enriched for mutated proteins comprising EGF/YES, Ubiquitination, and RPL26/5 nodes.

Conclusion: Our study produces a well annotated p53-dependent and p53-independent mutant proteome of a common melanoma cell line model. Coupled to the application of an integrated DNA and RNA variant detection platform (CLCbio) and software for identification of proteins (ProteinPilot), this pipeline can be used to detect high confident mutant proteins in cells.

General significance: This pipeline forms a blueprint for identifying mutated proteins in diseased cell systems.

INTRODUCTION

Next-generation genome sequencing technologies have revolutionized our understanding of the molecular nature of cancer [1] [2]. Parallel innovations in combinatorial chemistry, crystallography, high-throughput drug screening, transgenesis, and computational science have rapidly generated hundreds of promising targeted drug leads. However, despite this increased R&D, the number of effective drugs reaching the clinic is in steady decline [3]. A technical problem is the lack of robust age-dependent sporadic immune competent models of human cancer that predicts human toxicity and response [4]. Another major hurdle is that sporadic cancers are multi-gene diseases thus minimizing the likelihood of finding a common (set of) drug(s) to improve patient welfare.

Whole genome cancer sequencing has defined the strikingly patient-specific cancer bar code thus highlighting the unique genetic signature of any given tumour [5] [6]. However, the vast majority of anti-cancer medicines target wild-type proteins, although there are ever emerging successes in targeting mutated kinases with effective drug leads [7]. This presents an opportunity to develop precision, personalized therapeutics based on expressed, mutant proteins. Such mutant proteins could inform target pathway choice for the development of novel Biologics that target the mutated cancer landscape in a patient-specific manner [8]. Understanding the expression of mutant proteins could also form a platform for the development of mutant neoantigen based anti-cancer vaccines, that could be based on synthetic proteins [9], dendritic cells primed with neoantigens [10], nucleic acids such as RNA [11], or synthetic viral vectors [12]. However, the study of mutant proteomes is only in its infancy. The relative difficulty in this task is that building mutant proteomes requires integration of robust and user-friendly methods linking the fields of informatics, mass spectrometry, and cancer biology. This task is not necessarily trivial.

Computational methodologies as applied to the cancer research field are emerging as approaches to define the expressed, mutated genome. There are several challenges with optimizing platforms that integrate DNA sequencing, RNA sequencing, and mass spectrometric datasets [13]. One overall challenge is integrating these molecular data into one pipeline; and depends on the variant detection platform used for DNA sequencing; the algorithms for defining mutations with RNA sequencing datasets in exons, non-coding RNAs, and introns [14]. For example, the expression of intron encoded mutant peptides is almost completely unexplored at a systems biology level [15], as are cancer-specific RNA edits and tumour-specific spliced mRNAs that create cancer-associated polypeptide epitopes. In addition, different mass spectrometers, sample preparation and pre-fractionation methods, coupled to tumor heterogeneity, result in an incomplete understanding in the source and extent of expressed mutated proteins using cancer-specific DNA and/or RNA sequencing reference databases.

There are several types of platforms recently developed for integrating DNA, RNA and protein data integration. ‘Proteoformer’ produces a complete protein synthesis database that can be used to identify peptides with mass spectrometry through the use of ribosome profiling [16]. The limitation of this approach is that living cells are required to isolate bioactive ribosomes and this method might not be conducive to frozen tissues from the clinic. Methods have been established for automating spliced variants in cell systems which is especially powerful in cancer genomes where there might be DNA fusions, cancer-specific splicing, and trans-splicing [17] [18]. Spliced variant detection algorithms are always improving, especially those that capture the pathological, heterogeneous splicing specific to cancer cells [19]. Modification of R-packages iterates innovations in identifying expressed mutated genes [20]. Translation toolkits have been generated that aim to produce a theoretical total polypeptide space of a genome using RNAseq that captures

six-frame genome translation [21]. These examples highlight the types of several bespoke algorithms that generate cutting-edge information on the proteogenomic landscape. As the diversity in software and computational tools tend not to be benchmarked against each other, end-user compatibility, especially crossing different disciplines, can be limited or non-accessible.

In this report, we focus first on benchmarking two distinct DNA and RNA variant calling platforms towards identifying the mutant proteome landscape in a biological human cancer cell model. One of the variant detection platforms; *CLCbio*, is an integrated DNA and RNA variant identification software platform. This software has been used previously in variant detection using human cancer tissue [22] [23] although it was not benchmarked against more classic variant calling platforms. The utility of the *CLCbio* application is that it is a tool not requiring computational coding and can therefore open the door for life-scientists to identify tissue or cell-specific genomic variants relevant to the biological system of interest. Such an application for life-scientists that does not require coding already exists for research in the mass spectrometric field, such as *Proteome Discoverer* [24]. The most common coding-dependent genome analysis toolkits for DNA variant detection are platforms such as Varscan and Mutect [25]. In this current study, we not only benchmark both the *CLCbio* software and *Varscan2* platforms, we also use these data to create a mutant reference dataset to define mutated proteins in a key cancer cell model. We focus our biological question on asking an emerging topic in the cancer research field; how does inactivation of a major cancer-associated gene impact on the mutant proteome landscape? We answer this question by creating an isogenic melanoma cell panel with a p53 gene inactivation to define changes in the mutant proteome as function of p53 gene inactivation.

RESULTS

Using CLCbio and Varscan2 DNA variant detection platforms to develop a mutant genomic reference database using a human melanoma cancer cell line model.

The human melanoma A375 cell line has classically been used as a model to study regulation of wt-p53 activity in response to DNA damage [26] [27]. Next generation DNA sequencing of the A375 wt-p53 containing melanoma genome was performed in order to annotate its genome. This also produces a community resource that can be used to develop proteogenomics tools and pipelines for use in studying both mutated proteome expression and/or neoantigen production. The majority of next-generation data analysis using DNA variant detectors derived from *Varscan* or *Mutect* requires computational coding skills [25]. By contrast, the *CLCbio* platform that has been developed as an independent variant detection platform for life-scientists with plug-ins that do not require computational coding to define variants [22]. In this report, we benchmark both *CLCbio* and *Varscan2* as two independent variant detection platforms to define the overlap in their mutation detection and define their dual utility in creating a mutant genomic reference database for optimizing mutant peptide detection using mass spectrometry (Fig. 1).

Exome Sequencing of DNA derived from A375 cells was performed using Agilent V5+UTR Exome Capture Kit and 100 bp paired-end reads were acquired using a coverage of 100x. Paired fastq files were imported into the *CLC Biomedical Genomics Workbench 3*. Adaptor sequences and bases with low quality were trimmed, DNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through dbSNP databases to remove “common” germline variants. A total of 120,325 DNA variants were detected at a frequency of 5% or higher at the threshold used (Supplementary Table 1A). Filtering this list of tumour specific variants to a frequency of 40% or higher in order to capture the most dominant mutant alleles in the cell population, resulted in a total of 63,880 variant mutations detected (Supplementary Table 1B).

When the *CLCBio* variant calling platform was compared to *Varscan2* using the hg19 reference database, as well as the more recently updated hg38 reference genome, then 85,793 shared variants were detected (Fig. 2A). The *CLCbio* platform detected more variants than *Varscan2* at the thresholds used; 36,065 variants were unique to *CLCbio* using the hg38 reference genome; 17,824 variants were unique to *CLCBio* using the hg19 reference genome, and 3,180 variants were unique to *Varscan2* using the hg38 reference genome (Fig. 2A). Because *CLCbio* generally identified more variants than *Varscan2*, we focused on using this platform to optimize mutant peptide detection by mass spectrometry (Fig. 1).

Using CLCbio RNA variant detection platform to develop a mutant peptide reference database from a human melanoma cancer cell line

Of the filtered *CLCbio* DNA variants detected (Supplementary Table 1B), 20,419 were synonymous mutations within exons, 41,993 variants resided out with exons, and the 1,468 were non-synonymous mutations. This number of non-synonymous mutations is within the expected range of a tumour like melanoma which has a relatively high number of single nucleotide variants (Supplementary Table 2). The 1,468 protein-coding variants were derived from 884 genes, including single-nucleotide polymorphisms, in frame-insertions, and in-frame deletions. This list is very conservative and could be expanded by including DNA sequencing reads below the 40% threshold level (Supplementary Table 1B). Representative *CLCbio* browser views summarizing DNA variants are represented in Fig. 2B and 2C.

We next used shotgun RNAseq data derived from A375 cells to identify expressed mutated genes under conditions in which more liberal variant calling in the DNA variant files could be tolerated. RNAseq from A375 cells was performed using human total RNA, depleted of ribosomal RNA, followed by random priming to generate cDNA. From this template paired-end Illumina HiSeq2500 was used to generate approximately 20 million reads. Paired fastq files (available upon request) from RNAseq reads were imported into the *CLC Biomedical Genomics Workbench 3*. The RNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through the A375 cancer genome sequence where at least 2 mutant DNA reads were identified, then dbSNPs were removed, to identify mutated and expressed genes. A total of 18,341 expressed non-synonymous mRNA variants were identified with a cutoff of 5% of the total RNA reads defined as mutated (Supplementary Table 3).

Although DNA variant calling would typically rule out the annotation of 1 mutant DNA sequencing read, we also filtered the fastq DNAseq files (Supplementary Table 2) against fastq RNAseq data (Supplementary Table 3) with a stringent cutoff requiring at least 40 mutated RNA sequencing reads (e.g. relatively highly expressed mutant alleles) and filtered against a more liberal DNA variant cutoff of at least 1 mutant genomic DNA sequencing read. This produced 5,980 RNA variants including synonymous, non-synonymous, and non-exonic mutations (Supplementary Table 4). Upon filtering for non-synonymous variants, a list was generated composed of 1,418 non-synonymous highly expressed RNA variants encoded by 976 mutant genes (Supplementary Table 5). We then determined the overlap of the 1,468 *CLCbio* derived DNA variants using the stringent DNA variant calling (Supplementary Table 2) to the 1,418 expressed mutated RNA variants identified using liberal DNA variant calling but requiring high levels of mutated mRNA reads; e.g. a highly expressed mutated gene (Supplementary Table 5).

The first thing to note is that the list of highly expressed mutant genes selected based on the numbers of RNA variant reads is highly divergent from the list of mutant genes selected based on

the number of DNA sequencing reads (Fig. 2D). Only 107 variant genes are shared in this subset. One example of an overlap between the DNA and RNA variant calling cutoff highlight the expressed mutant RNA derived from a mutant gene is the *gpatch4* gene (Fig. 2E). This gene is homozygous mutant (Supplementary Table 1A and B). A total of 877 out of 985 genes with expressed RNA variants are not present in the original DNA variant list (Supplementary Table 6). This produces a mutant gene expression rate of 107/774 or 12.1%. The true value will be higher than 12.1% since we removed RNA variants that exhibited lower than 40 reads. This mutant gene expression frequency is in lower range of ~30% previously published; previous studies have shown that 36% of validated somatic SNVs were observed in the transcriptome sequence when RNAseq data was compared with the genomes/exomes data in breast cancer [28] and similar proportions were also observed in a lymphoma study in which 137 somatic mutations were expressed in RNAseq, out of 329 total somatic mutations [29].

Mutated protein identification using the A375 DNA genomic reference database

We next aimed to define the extent to which the DNAseq or DNA+RNAseq reference databases could be used to identify mutated peptides by mass spectrometry. Proliferating A375 cells were lysed and protein was processed using the FASP (Filter-aided sample preparation) method [30]. Measured spectra were processed in ProteinPilot 4.5 search engine where a Swiss-prot and TrEMBL reference search database (Supplementary Table 7) was used (as described below) [31]. This produced a file of 949 wild-type proteins identified at FDR<1%. In order to determine whether any of these wild-type identified proteins are mutated, we next filtered the 1,468 non-synonymous DNA variant set (Supplementary Table 2) with the 949 detectable protein set (Supplementary Table 7) to generate a list of 42 potential mutant polypeptide sequences (Table 1). Only one high confidence mutated tryptic peptide covering a sequence of mutated protein from this group was identified. The peptide was derived from the ribosomal protein rpl14 (Fig. 3A and 3B). The mutant peptide spectrum and the fragmentation summary were exported from the ProteinPilot 4.5 search engine (Fig. 3C and 3F). The wt-rpl14 peptide covering the same position in protein sequence was also observed (Fig. 3D and 3E) which is consistent with the heterozygous mutation identified by DNA sequencing (Fig. 3A).

Limitations of mass spectrometry to identify mutated protein sequences.

There could be several reasons why the majority of these mutated proteins might not be detectable using mass spectrometry. First, as 18 out of the 42 proteins from this group have a homozygous gene mutation (Supplementary Table 1 and Table 1), these 18 expressed proteins presumably are mutated. However, there were no mutated tryptic peptides derived from any of these 18 proteins. This highlights a general difficulty in relying on mass spectrometry to confirm the expression of mutated proteins. The number of identified peptides covering any one protein is rarely “100%”. For example, the tryptic coverage of rpl14 highlights just this problem, as only two out of over 16 theoretical tryptic peptides (over 8 aminoacids of greater in length) could be detected (Fig. 3B). Serendipitously, one of these two peptides covered the mutant region (Fig. 3C and 3F). Other reasons for the absence of mutant peptide detection is that the mutant tryptic peptide could be relatively small (e.g. less than six amino acids in length) and therefore difficult to confidently match to corresponding MS/MS spectrum. For example, the protein LRCC59 has an in-frame triplet nucleotide insertion resulting in a Q amino acid insertion (Lys137_Pro138insGln) resulting in the tryptic penta-peptide QPFPK. It is difficult to identify this mutated peptide, unambiguously, as being derived from LRCC59 as there are several proteins in the human proteome that could yield this amino acid sequence after trypsinization at the sequence KQPFPK. The inability to detect such mutant tryptic peptides from any of the 18 proteins derived from homozygous mutant genes provides a measure of the false negative discovery frequency and

highlights the need to include different proteases in sample preparation or derive a larger pool of tryptic peptides from which to search for mutated peptides. Nevertheless, users of this A375 cell model as a resource might want to consider these 18 proteins (with no detectable mutant tryptic peptides) expressed from homozygous mutated genes to be *bona-fide* mutant proteins.

Mutated protein identification using mass spectrometry with the A375 mutant RNA reference database

As mutant peptides were not detectable to a high degree using the DNAseq-only files, we focused on using the RNAseq files stratified by 40 or more mutated mRNA reads derived from wt-p53 A375 cells (Supplementary Table 5) to create a mutant search database. We also next initiated an additional pipeline approach that was used to increase number of detectable proteins by pre-fractionating the peptides using an orthogonal LC method (Fig. 1). The orthogonal approach incorporated a reverse phase high pH acetonitrile gradient generating ten fractions that were infused for peptide identification using a TripleTOF 5600+ mass spectrometer. Employing peptide pre-separation step increased the coverage of total wild-type peptides to over 35,000 (FDR<1%) with a coverage of over 4,500 wild-type proteins (FDR<1%) (Supplementary Table 8).

Integrating the RNAseq derived mutant search database (Supplementary table 8) to the 2D LC-MS/MS data increased the number of identified mutant peptides to 193 (Supplementary Table 9). Although the ProteinPilot 4.5 search engine determines the confidence of identified mutated peptides, we needed to manually inspect the spectra in each of these 193 proteins to produce a list of mutated peptides that were identified based on y and b fragment ions covering the mutated amino acid in the mutated peptide sequence. Applying this procedure excludes the possibility that we identified wild-type peptides and narrowed the list down to 60 verified mutated proteins (Fig. 4).

Methodologies for validation of mutated peptides in the wt-p53 A375 cell model.

Mutated peptide identification in A375 fractions was further examined using targeted mass spectrometry in pseudo-selected reaction monitoring mode (pseudo-SRM) on TripleTOF 5600+ mass spectrometer. Ten isotope labeled mutant tryptic peptides were acquired (Fig. 5B) to validate ten of the 60 mutant tryptic peptides identified in data-dependent mode (Fig. 5A). We optimized isotope labelled peptide spike-in into A375 lysates (Fig. 5C and 5D) to yield optimal product ion intensity. Comparing the retention times and product ion intensity patterns in product ion chromatograms of naturally occurring intrinsic peptide and isotope labelled peptide enabled us to determine whether the naturally occurring peptide is present. All ten of the mutant peptides were successfully validated by this methodology.

As an example, verification of two mutated peptides is shown in Fig. 6A and 6B. Fragmentation evidences described in Fig. 6A and 6B show all possible product ions that represent two selected high confidence mutated peptides (peptide confidence > 99%) identified in 2D LC-MS/MS data (VSGSPEQAVEENLSSYFLDR and IIPTVLMTEDIK peptides). Only underlined product ion masses represent identified product ions that confirm the presence of amino acid mutation in the peptide sequence. The fragmentation evidence of the VSGSPEQAVEENLSSYFLDR mutated peptide highlights 11 product ions confirming the S to F mutation, whilst the fragmentation evidence of the IIPTVLMTEDIK mutated peptide shows only 3 product ions confirming the A to P mutation**.** The fragmentation evidences in Fig. 6A and 6B clearly show that the probability of an amino acid mutation is not reflected in the peptide confidence determined in ProteinPilot 4.5 software. Therefore, it is important to evaluate the spectra/fragmentation evidence to accurately define mutant status. An example SRM validation of a mutant peptide (SIITYVSSLYDTMPR) with heavy

isotope labelled reference peptide and its light naturally occurring variant is shown in Fig. 6C. All results from SRM validation of mutant peptides using SRM are summarized in Supplementary Table 10 and Fig. 6D. Taken together Fig. 6 summarizes all steps, that were taken to validate any of the selected mutated peptides.

Mutated proteomics: Creating a p53-null A375 melanoma cell line as an isogenic model system to define p53-dependence in mutated cancer genome protein expression patterns.
We finally aimed to use this optimized DNA and RNAseq variant detection pipeline (Fig. 1) to ask a key biological question; how does inactivation of the tumour suppressor p53 gene impact on the mutant proteome landscape? The human melanoma A375 cell line has classically been used as a model to study regulation of wt-p53 activity in response to DNA damage [26] [27]. We thus focused on using CRISP-R gene editing to develop an A375 cell line with an isogenic p53-null status. The guide RNA encoding a targeting sequence near the ARG175 codon in the p53 gene (Fig. 7A) was transfected into cells, single cells were isolated following cell sorting, and individual clones were selected based on absence of p53 induction after X-irradiation (data not shown). Nine independent p53-edited clones were obtained (data not shown). One A375 p53-null cell clone was taken forward with the edits as indicated in Fig. 7B and 7D. The sequences across the breakpoint in edited p53 alleles (Fig. 7) results in a stop codon (Fig. 7C and 7E). Immunoblotting of lysates using the N-terminal epitope antibody DO-1 confirmed that the A375 p53-null cells do not express p53 protein, nor do the HCT116 p53-null cells (Fig 7F). In addition, MHC Class I protein (HLA-B allele) was also determined to be elevated in the p53-null cells (Fig 7G). We also observed elevated HLA-A and HLA-C alleles in p53-null cells (data not shown), further highlighting the utility of this p53 –null cell as a tool to study in the future how p53 status impacts upon mutated protein expression as well as mutant peptide presentation by the MHC Class I system.

SWATH (Sequential Windowed Acquisition of all Theoretical Mass Spectra) was used as a complementary methodology to further determine the presence of target mutated peptides and to evaluate an effect of p53 inactivation on mutated protein levels. We first set-up large scale SWATH quantitation on TripleTOF 5600+ mass spectrometer and focused towards two mutated proteins that were previously verified and validated in wt-p53 A375 cells; PYGB and PLEC (Fig. 8). We extracted product ion chromatograms corresponding to these mutated peptides using a product ions m/z's included in a spectral library derived from data-dependent measurement of same sample. We found high correlation between product ion intensity pattern in extracted product ion chromatograms and the product ion intensities in spectral library for both peptides (Fig. 8A, 8B, 8D, 8E). Following quantification of the PYGB mutant peptide using three technical replicates obtained from A375 p53-null cells and A375 p53 wild-type cells shows an up-regulation of the PYGB (LIINLVTSIGDVVNHDPVVGDR) mutated tryptic peptide in A375 null cells. However, the quantification we observed based on the set of automatically intensity based selected product ions does not specifically represent the mutated peptide, and as such we cannot rule out the possibility that the quantitative data are derived from wild-type tryptic peptide. Therefore, as a robust SWATH pipeline we would again recommend to carefully select product ions referring uniquely to the mutation in a peptide sequence. In case of PYGB mutated peptide (LIINLVTSIGDVVNHDPVVGDR) it would be ions encompassing b4 – b21 and y19 – y21. The set of automatically selected product ions from spectral library contains y6, y16, y17, y9 (Fig. 8A). Therefore, SWATH quantitation of peptide LIINLVTSIGDVVNHDPVVGDR does not uniquely refer to the mutated peptide form and could be biased by changes in wild-type form.

We next extracted quantitative SWATH data for mutated peptide from PLEC (SIITYVSSLYDTMPR) that was also successfully verified and validated (Fig. 8D). In this example, we extracted

quantitative data referring to product ions that uniquely involve mutated amino acid in the product ion sequence. The mutated peptide SIITYVSSLYDTMPR must be quantitated by any product ions encompassing b12 – b15 and y4 – y15. The set of automatically selected product ions from spectral library for quantitation of SIITYVSSLYDTMPR 3+ contains y4, y5, y6, y7 (Fig. 8D). Therefore, quantitation of the mutated peptide refers uniquely to mutated peptide form. The corresponding spectral library evidence and product ion intensity rank in product ion chromatogram are highlighted (Fig. 8D and 8E) along with the quantitation of mutated peptides in three technical replicates of A375 p53-null cells and A375 p53 wild-type cells (Fig. 8F). As a control, SWATH quantitation of a control peptide VAPEEHPVLLTEAPLNPK (ACTB) was used to evaluate quantitative differences as a consequence of deleting the p53 gene in A375 cell line, differences in sample loading, and the possible effects of sample preparation on MS analysis (Fig. 8G).

Differences in mutant protein expression in wt-p53 and p53-null cells.
We finally determined the major differences in mutant peptides enriched between wt-p53 and p53-null cells using *CLCBio* and *Varscan* variant detection platforms. Parameters were set requiring at least 10 mutant mRNA variants and 1 mutant DNA variant. Using *CLCbio*, there were 190 mutant peptides detected enriched in wt-p53 cells, with 187 mutated peptides enriched in the p53-null cells (Fig. 9A). The largest number of mutant peptides were detected using *CLCbio* when compared to Varscan (Fig. 9B and 9C). Using STRING protein annotation, we evaluated the dominating mutant protein networks defined (Fig. 10 and 11).

In cells containing wt-p53 there were dominating peptides that could be used to highlight mutated protein networks centered on CDC5L (Fig. 11, center). The CDC5L network in turn is linked to a mutant protein splicing network composed of SRSF7, SF35B, SLU7, CSTF3, and YBX1. In turn these networks also connect to mutant proteins in DNA repair pathways including POLR1b, ERCC2, FANCD2, XRCC6, and others (Fig. 10). By comparison, the STRING analysis in p53-null cells highlighted a different dominating mutant proteome. This consisted of a ribosomal mutant protein network including RPL5 and RPL26L (Fig. 11). This node was in turn connected to a mutant ubiquitin protein node including USP7, UBE2V2, USP9X, and MRE11. An independent node composed of mutant proteins was present in kinase signalling including EGF, YES1, PPP2CB, CSK, and PPP2R5C (Fig. 11). Together, these data highlight that inactivation of the p53 gene can switch the expression of distinct mutant protein signalling nodes encoded by a cancer genome. This in part, can shed light on how tumour cells adapt to gene mutation by changing the expression of mutant proteins that comprise specific signalling pathways.

DISCUSSION

The identification of mutated proteins in human cancer cells can assist in defining expressed, mutated oncogenic signalling landscapes as well as facilitating the identification of potential neoantigen vaccine ligands. Most often, proteomics studies using cancer cell lines uses a normal reference proteome and under-estimates mutated protein signalling functions. We report on an optimized pipeline for identifying tumour variants in A375 cells using; (i) the *CLCBio* integrated DNA and RNA variant calling platform; (ii) the incorporation of RNAseq to stratify highly expressed variants; (iii) the use of 2D LC-MS/MS to identify potentially mutated peptide sequences in a tumor cell line; and (iv) the use of manual spectral annotation and SRM to estimate a false discovery rate of mutant peptides using LC-MS/MS. Our pipeline has identified high confidence list of protein mutations in the A375 cell line and its p53-null derivative (Supplementary Tables 13 and 14) by stratifying genetic mutations based on high levels of mutant mRNA expression and using

shotgun mass spectrometry to identify mutated peptides. This underestimates the extent of mutant protein expression in the A375 cell because of; (i) incomplete tryptic coverage of any one given protein; (ii) we only included very abundant mutant mRNA species (with relatively high numbers of reads; i.e. >40) to create a reference database of relatively highly expressed mutant genes; and (iii) stringent manual annotation of mutated peptides identified in the mass spectra eliminates some likely mutant peptides from the dataset.

We focused this study on using a human melanoma has emerged as a cancer type with one of the highest rates of single-nucleotide variation in a cancer genome and this tumour type is expected to form important models to define mutated protein expression networks [32]. As a result, melanoma patients can benefit from cancer-specific immuno therapeutics that exploit this high rate of mutated protein production [33]. Understanding how cancer associated genes impact on steady-state mutant protein expression levels and ultimately neoantigen assembly into the MHC Class I pathway has thus recently gained more relevance. Thus, it is important to begin to develop isogenic cancer cell models with specific cancer gene mutations to accelerate our understanding of how mutant protein production and re-wired mutated signalling proteins are enhanced in cancer cells. This could facilitate developing new therapeutics that exploit mutant protein or mutated pathway expression. We focused here on applying these methods to the question of how p53 gene inactivation can impact on mutant protein expression in an isogenic melanoma cell model.

P53 protein has been termed the guardian of the genome [34]. The p53 gene is one of the most frequently mutated in the vast majority of cancers [35] leading to loss of wt-function and enhanced genome instability [36]. It is therefore interesting that the majority of human melanomas retain wt-p53 alleles [32], possibly because the selections pressures driving the survival of cells with a mutant p53 gene is reduced due to frequent inactivation of the p53 regulatory *cdkn2a* locus in melanoma. Oesophageal adenocarcinoma also has very high rate of single nucleotide variation [37] but also a high rate of both p53 gene mutation and *cdkn2a* mutation [38]. These data might suggest there could be two distinct pathways that drive a high rate of single nucleotide variation in a cancer genome via attenuation of wt-p53 function (as in melanoma) or mutation in the p53 gene (as in oesophageal adenocarcinoma). We focus in this report on generating an isogenic wt and p53-null cell panel using CRISP-R/CAS9 gene editing methodology to define how loss of p53 can impact on a mutant proteomic landscape. Integrating genomics, RNA expression, and mass spectrometric data produces a baseline mutant protein library in melanoma that can be used as a community resource to facilitate interrogation of signal transduction pathways in this model. A similar cancer cell model has been developed in oesophageal adenocarcinoma that has the features of high rate of single nucleotide variation, but which has mutations in the p53 tumour suppressor gene [39]. Using our optimized pipelines, we define specific sets of mutant proteins that are differentially expressed as a result of p53 gene status. These data indicate that loss of a tumour suppressor gene such a p53 can begin to switch expression of the mutated protein landscape in a tumour cell (Fig. 10 and 11).

One impact of this pipeline will be in the future identification of tumour-specific mutated proteins that can be processed by the MHC Class I pathway; neoantigens. Within this list of mutant proteins detected in the A375 cell line, we were also able to identify some trimmed peptides that have a predicted high affinity for MHC class I peptides as defined using netMHC 4.0 [40], based on our isotyping that A375 cells have the MHC Class I alleles, HLA-A*-3:01 and HLA-B*07:02. One of these includes the gene HIST3H2A, which has a V108L mutation. The predicted affinity of the wt-

10mer LPNIQAVLL for HLA-B*07:02 is 518.5 nM and the predicted affinity of the mutant-10mer LPNIQAVVL for HLA-B*07:02 is 87.8 nM. Additionally, SRP14, with a P124A mutation resulted in a mutant 10mer APAAAATAA peptide with a predicted affinity of 17.9 nM for HLA-A*03:01 whilst the wt-peptide APAAAATAP peptide with a predicted affinity of 641.5 nM for HLA-A*03:01. Such datasets will provide neoantigen models to study MHC Class I peptide flux in this A375 cell line.

A second impact of this pipeline will be on mutant proteomics studies. We have used a gene editing tool (CRISP-R) to begin to ask a biological question on how loss of p53 can impact on the mutant protein landscape. Pathway annotation using STRING provides evidence for mutated protein expression in pathways including adhesion, ubiquitination, metabolism, and DNA repair. Mutant protein expression in a cancer cell line provides another way of thinking about 'proteomics' compared to its usual application which is examining protein expression in a cancer cell line using a 'normal' reference protein database. When screening the mass spectral data from A375 cells and the p53-null A375 cells against the mutant genomic reference databases, we identified a relatively high overlap in the proteins identified between the two cell lines (Fig. 9A). This might not be unexpected since p53 is stress activated. On the other hand, the difference in the proteome between the two cell lines (Fig. 9A) also highlights the fact that inactivation of the p53 gene did give rise to spontaneous changes in some mutated proteins without further selection pressures. Future studies could examine how loss of p53 in this isogenic model impacts on radiation or GAS/STING-dependent mutant protein signalling. Together, our study produces a well annotated mutant proteome of this A375 cell line model, coupled to the application of an integrated variant detection platform (*CLCbio*), that can be used to detect high confident mutant proteins cells. This can facilitate the use of this isogenic model as a resource to identify the p53-dependence on the mutant proteome and normal proteome landscape.

Acknowledgement: The work was partially supported by the European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868) and MH CZ - DRO (MMCI, 00209805), JF, PM and FZK were supported by AZV NV18-03-00339; the BBSRC RASOR consortium (BB/C511599/1; United Kingdom); the Wellcome Trust (grant number 094417/Z/10/Z); The International Centre for Cancer Vaccine Science project carried out within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund; the National Natural Science Foundation of China (Grant. 81701883) and China Scholarship Council.

Materials and Methods

P53 knockout using the CRISPR/CAS9 system.

The p53-specific gRNA sequence was 5'-CTGAGCAGCGCTCATGGTGGNGG-3', which was designed by Applied StemCell, Inc. The gRNA was cloned into pBT-U6-CAS9-2A-GFP expression vector from Applied StemCell, Inc. The P53 knockout in A375 cell line was performed as described before with minor alterations [41]. Briefly, $3x10^5$/well A375 cells [27] were seeded in 6-wells plates. 24 hours later, cells were transfected with pBT-U6-CAS9-2A-GFP expression vector using Attractene Transfection Reagent (QIAGEN, UK). 48 hours later, mutations were tested using a Surveyor Mutation Detection Kit (Integrated DNA Technologies, USA) and GFP positive cells were sorted and collected by BD FACSCanto II (BD Bioscience, USA). GFP positive cells were seeded in 96 wells plates 1 cell/well for colony formation. After 2 weeks, all colonies were collected and tested by western blot using the in-house developed DO-1 monoclonal antibody to demonstrate loss of p53 protein (Fig 7F) and sequencing to validate p53 gene editing in the A375 cell lines. The HLA-B antibody was from Thermofisher (PA5-35345). The loading control for immunoblotting (Fig 7F) was an in-house antibody developed to PCNA. The antibody used as a loading control (Fig 7G) was

mouse monoclonal anti-β-actin (Sigma). The HCT116 wt and p53-null cells were a gift of Dr. B Vogelstein (Johns Hopkins University, USA).

CLCbio variant calling.

Next generation DNA sequencing of the A375 wt-p53 containing melanoma cell line was performed in order to annotate its genome. Exome Sequencing was performed using Agilent V5+UTR Exome Capture Kit (75Mb) and 100 bp paired-end reads were acquired using a coverage of 100x. Paired fastq files from the A375 cell line (available upon request) from DNA-exome libraries were imported into the *CLC Biomedical Genomics Workbench 3*. Adaptor sequences and bases with low quality were trimmed, DNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through dbSNP databases to remove "common" germline variants. A total of 120,325 tumour specific variants were detected at a frequency of 5% or higher when the threshold was set at calling variants detected in at least two variant DNA sequencing reads in the exome data from A375 cells (Supplementary Table 1A). RNAseq from A375 cells was performed using human total RNA, depleted of ribosomal RNA, followed by random priming to generate cDNA. From this template paired-end Illumina HiSeq2500 was used to generate approximately 20 million reads. Paired fastq files (available upon request) from RNAseq reads were imported into the *CLC Biomedical Genomics Workbench 3*. The RNA sequencing reads were mapped to the human reference genome hg19, and sequences were filtered through the A375 cancer genome sequence where at least 2 mutant DNA reads were identified, then dbSNPs were removed, to identify mutated and expressed genes. A total of 18,341 expressed non-synonymous mRNA variants were identified with a cutoff of 5% of the total RNA reads defined as mutated (Supplementary Table 3).

Varscan2 variant calling.

The same paired fastq files from the A375 cell line (as used for *CLCbio* analysis, above) from DNA-exome libraries were used for following analysis pipeline. Variant calling was performed using Varscan, adaptor sequences and bases with low quality (QualityScore < 30) were trimmed by FASTX-Toolkit version 0.0.14 (Retrieved from http://hannonlab.cshl.edu /fastx_toolkit), DNA sequencing reads were mapped to the human reference genome hg19 and hg38 (by TopHat2 [42]) (See Fig. 1), and results were imported into VarScan2 [43]. Sequences were filtered through dbSNP databases to remove "common" germline variants. A total of 101,072 tumor specific variants (coverage 1 DNA mutation) and 10,545 non-synonymous mutations were detected using hg38 (coverage 1 DNA mutation; See Supplementary Table 11). A total of 17,822 tumor specific variants were obtained from RNA seq (using a minimal cutoff of 10 mutated RNA reads) and 2,662 were classified as a non-synonymous RNA mutation (using a minimal cutoff of 10 mutated RNA reads). A total of 5,590 tumor specific variants were obtained from RNA seq (using a minimal cutoff of 40 mutated RNA reads, data not shown) and 1,007 were classified as non-synonymous RNA mutations (using a minimal cutoff of 40 mutated RNA reads). The number of 2,461 tumor specific variants from the wt-p53 A375 cell line (1 DNA mutation and at least 10 mutant RNA reads) and 945 tumor specific variants from the p53-null cell line (1 DNA mutation and at least 10 mutant RNA reads), were used as the input file for reference database to identify mutant peptides using mass spectrometry Supplementary Table 12. A comparison of the DNA mutations detected using Varscan and CLCbio are summarized in Fig. 2 and the number of mutated peptides detected, using either Varscan or CLCbio driven analysis are summarized in Fig. 9.

Peptide sample preparation for MS.

Cells were plated and grown on five 10 cm Petri dishes to 80% confluence. Cells were harvested into lysis buffer composed from 8 M urea in 0.1 M Tris/HCl pH 8.5 (urea buffer). Cell lysis was

facilitated using needle sonication in three 4 second cycles and protein concentration was determined using RC-DC assay (Bio-Rad, USA). An aliquot corresponding to 200 µg of protein was digested to peptides using Filter-aided sample preparation (FASP) [30]. Briefly, cell lysate was mixed with 200 µl of urea buffer and added to centrifugation filter unit Vivacon 500 with 10 kDa cut-off (Sartorius Stedim Biotech, Germany) followed by centrifugation (15000 g/ 20 min/ RT). 16.7 mM TCEP in urea buffer was added to filter unit to reduce disulphide bridges in protein. Reduction was done on thermomixer (600 rpm/ 30 min/ 37˚C) followed by centrifugation (15000 g/ 20 min/ RT). Sample alkylation was performed in the darkness for 20 min at RT with 300 mM Iodoacetamide in urea buffer followed by centrifugation (15000 g/ 20 min/ RT). Buffer was exchanged to 100 mM NH_4HCO_3 using 3 washes to enable efficient trypsin digestion of samples. Proteins were digested in 100 µl of 50 mM NH_4HCO_3 buffer where 1 µg of trypsin was added per each 33 µg of protein to be digested. The samples were mixed with digestion buffer at 600 rpm on thermomixer for 1 min before incubation for 18 h in a wet chamber, 37˚C. Peptides were eluted from the filter by centrifugation (15000 g/ 20 min/ RT). To increase peptide recovery 0.5 M NaCl was added followed by centrifugation (15000 g/ 20 min/ RT). Peptide samples for direct MS analysis were desalted on Micro SpinColumns C18 (Harvard Apparatus, USA). First, C18 columns were conditioned twice with 100% acetonitrile (AcN)/ 0.1% formic acid (FA), centrifuged at (120 g/ 2 min/ RT), and washed with 0.1% FA followed by centrifugation (200 g/ 2 min/ RT). The columns were hydrated for 15 min in 0.1% FA, centrifuged (200 g/ 2 min/ RT). Samples were loaded to columns and centrifuged (550 g/ 2 min/ RT). Next, the columns were washed thrice with 0.1% FA. Peptide elution was done using 50% AcN/ 0.1% FA in water, 80% AcN/ 0.1% FA in water and 100% AcN/ 0.1% FA. The peptide eluates were evaporated using a SpeedVac and dissolved in 5% AcN/ 0.05% trifluoroacetic acid (TFA) in water. Concentration of peptides was determined in each sample on NanoDrop 2000 (Thermo Scientific, USA) at 220 nm and 280 nm prior MS analysis.

RP-basic fractionation using spin columns.

FASP digested peptide samples for *RP-basic fractionation* were separated in basic pH (pH 10) on Macro SpinColumns C18 (Harvard Apparatus, USA) packed with C18 sorbent. First, mobile phase A composed from 10 mM ammonium formate (AF) in water pH 10 and mobile phase B composed from 10 mM AF in 90% AcN pH 10 were prepared [44]. C18 columns were conditioned twice with mobile phase A followed by centrifugation at (200 g/ 2 min/ RT). Columns were then washed with mobile phase B followed by centrifugation (300 g/ 2 min/ RT). Columns were hydrated for 15 min using mobile phase A, centrifuged (300 g/ 2 min/ RT). Peptide samples were added to hydrated columns and centrifuged (650 g/ 2 min/ RT). Next, the columns were washed thrice with mobile phase A. Peptide separation into 11 fractions was done using a step gradient composed from 5% B + 95% A, 9% B + 91% A, 13% B + 87% A, 17% B + 83% A, 21% B + 79% A, 25% B + 75% A, 35% B + 65% A, 50% B + 50% A, 80% B + 20% A and 100% B. Ten fractions were evaporated using a SpeedVac concentrator. Each fraction was dissolved in 100 µl of 50% methanol and then evaporated in SpeedVac concentrator. This step was repeated three times to get rid of volatile salts. Dried samples were dissolved in 20 µl of 5% AcN / 0.05% TFA. Concentration of peptides was determined in each sample on NanoDrop 2000 (Thermo Scientific, USA) at 220 nm and 280 nm prior MS analysis.

Spiking sample with reference heavy peptides.

JPT synthetic heavy reference peptides (JPT, Germany) derived from 10 mutant peptide candidates were ordered with isotopically labelled amino acids. A content of the vial containing heavy reference peptide was dissolved in 5% AcN + 0.05% TFA in water to prepare 1.43 nmol/µl stock solution. Next, LOD of each peptide was determined (data not shown). A pool representing all 10 reference peptides was prepared and spiked into each peptide fraction (A375 peptide

concentration and injection volume was set based on crude Nanodrop measurement A_{220} and A_{280}) in order to load onto the column an identical amount of reference peptide corresponding to at least 10 times LOD. Three µl of this mixture were injected onto nano-LC-MS/MS to perform pseudo-SRM analyses in analytical triplicates.

LC setup for analysis of fractions.

Eksigent Ekspert nanoLC 400 system (AB-SCIEX, USA) nano-LC system was used for peptide concentration and separation. Peptides were concentrated and desalted on a cartridge trap column (300 µm i.d. × 5 mm) packed with C18 PepMap100 sorbent with 5 µm particle size (Thermo Fisher Scientific, MA, USA). Peptides were 10 min washed using 0.05% TFA in 5% AcN in water. Separation was performed on a 25 cm fused-silica emitter column with 75 µm inner diameter (New Objective, USA), packed in-house with ProntoSIL C18 AQ 3 µm beads (Bischoff Analysentechnik GmbH, Germany). LC solvents were composed from 0.1% FA in water (solvent A) and 0.1% FA in AcN (solvent B). Sample was eluted in a 120 min gradient starting at 5% B up to 40% B with a flow rate 300 nL/min in DDA and SWATH experiments while 61 min gradient starting at 5% B up to 40% B with a flow rate 300 nL/min was used in pseudo-SRM experiments. Peptides eluting from column were ionized in nano-electrospray and entered mass spectrometer.

Shotgun MS/MS (DDA).

Mass spectrometer TripleTOF 5600+ (AB-SCIEX, Toronto, Canada) operated in data dependent mode. Each cycle was accompanied with fragmentation of top 20 most intense precursor ions. Exclusion time was set to 12 seconds. Minimum precursor ion intensity was set to 50 cps and 100 milliseconds accumulation time per precursor. Shotgun were searched using ProteinPilot 4.5 (AB-SCIEX, Canada) against custom built human mutant proteome reference database derived from human reference database (Uniprot+Swissprot 2016_2) and against wild-type human reference proteome database (Uniprot+Swissprot 2016_2). Search parameters were set as follows: trypsin protease, carbamidomethyl (C) (fixed). Protein FDR was determined by searching MS/MS data against decoy databases.

Building-up custom search library with mutant protein sequences.

MS/MS data from fractions were inspected for presence of MS/MS spectra corresponding to genes identified as high confidence RNA and DNA variant hits. Amino acid sequences corresponding to these genes were listed in a mutant search database where a mutant position in a protein was inserted based on genomic data. Mutant search database was assembled in FASTA format. Corresponding wild-type FASTA sequences of protein forms were downloaded from the human (2016_2) reference database. A custom mutant FASTA file containing mutant and WT forms of proteins was created in a text editor and subsequently imported into ProteinPilot 4.5. Quality of mutant IDs was inspected in ProteinPilot 4.5. (AB-SCIEX, USA), focusing mainly to a peptide covering mutation position in a sequence. Spectral evidence of high and mid confidence mutant peptides (peptide confidence > 95% and between > 50% and < 95%) was manually inspected. In mutant peptides we focused to corresponding product ions directly proving a shift in mass as a consequence of mutation. A protein referred as “mutant” had covered potential mutation position by high or mid confidence peptides bearing substitution, insertion, deletion of amino acid in its sequences. On the other hand, a protein referred as “wild-type” has a place of potential mutation covered by wild-type sequence of a high or mid confidence peptide. Identification of mutant or wild-type form was accomplished by Uniprot BLAST of peptide sequence against Human reference proteome Uniprot 2016_2 to prove its proteotypicity.

SWATH-MS.

SWATH method for label free quantification was developed according to previously published methods [45] [46] [47]. TripleTOF 5600+ (AB-SCIEX, Canada) operated in high sensitivity positive polarity SWATH mode. Effective precursor range was selected from 400 amu up to 1200 amu and the cycle time was set to 3.5 seconds. An optimal SWATH width was 13 Da including 1 Da overlap resulting in a method with 67 SWATH windows. Accumulation time per SWATH was 50 milliseconds. Product ion range was scanned from 360 amu up to 1360 amu and rolling collision energy was used with collision energy spread (CES) of 15 mV.

Building-up wild-type spectral library and SWATH quantification of wild-type proteins.
PeakView software 1.2.0.3 (AB-SCIEX, Canada) was used to index 1045 proteins FDR 1% in a spectral library. Retention time window was set to 2.5 min to the left and 2.5 min from expected retention time. Protein quantitation was based on extracting peak areas for four peptides per protein and 6 product ions per each peptide. Extracted quantitative data from three technical replicates were statistically evaluated in MarkerView 1.2.1.1 (AB-SCIEX, Canada). Pairwise T-test was performed to determine protein fold changes and *P* values of fold change for all proteins listed in spectral library.

Building up a mutant spectral library and SWATH quantification of mutant proteins.
Results from mutant database MS/MS search were imported into Skyline-daily (64-bit) software version 2.6.1.6899 (MacCoss Lab, WA, USA) where the mutant spectral library was generated. FASTA sequences of 10 candidate mutant proteins listed in Supplementary Table 10 were imported into Skyline software. Results from mutant peptides are shown in Supplementary Table 10. An example of mutant peptides from RPL14 and PLEC is shown in Fig. 8. Peptide settings were as follows: digestion – trypsin, 1 missed cleavage was permitted, length of peptide was in range from 7 – 25 amino acids, 25 amino acids from N-terminus were excluded, carbamidomethyl was set as structural (fixed) modification. Transition settings were as follows: precursor charge was 2+, 3+ or 4+, fragment charge 1+, fragment series y or b, product ions, fragment ions from 3 up to last, including the N-terminal fragment to proline. After data extraction, start and end points of each peak in extracted product ion chromatogram were inspected manually. We checked co-elution of selected product ions, product ion ranks according to the spectral library, retention times across replicates, and data acquisition modes to confirm peak identity. Peak areas corresponding to mutant peptides were evaluated and visualized in Skyline software.

Pseudo-SRM analysis.
The TripleTOF 5600+ (AB-SCIEX, Canada) was operated in high sensitivity positive mode. Each cycle involved one TOF-MS scan with 250 ms accumulation time and 38 product ion scans of 70 milliseconds accumulation time per precursor. Total cycle time was 3.0 seconds. Product ions were scanned in a range from 100 to 1800 amu. Pseudo-SRM data were analysed in *Skyline*. Sequences of ten mutated peptide candidates were imported into the *Skyline* version 2.6.1.6899 (MacCoss Lab, WA, USA) with the following settings: Peptide settings were as follows: digestion – trypsin, 1 missed cleavages were permitted, length of peptide was in range from 7 – 25 amino acids, 0 amino acids from N-terminus were excluded, carbamidomethyl was set as structural (fixed) modification, heavy isotope peptide labeling 13C 15N on lysine, arginine and leucine was permitted. Transition settings were as follows: precursor charge was 2+, 3+, fragment charge 1+, fragment series y or b, product ions, fragment ions from 3 up to last, including the N-terminal fragment to proline. After data extraction, start and end points of each peak in extracted product ion chromatogram were inspected manually. We checked co-elution of selected product ions, product ion-rank according to the spectral library, retention times across replicates, and data acquisition modes to confirm peak identity. Peak areas corresponding to mutant peptides were evaluated and visualized in

Skyline software. dot-product (dotp) correlation for the ratio of the observed SRM peak intensities of a peptide in a specific biological matrix as correlation of observed SRM peak intensities of a peptide in a specific biological matrix versus the reference isotope labelled peptide were calculated. Peptides considered as high-quality validated hits showed good signal and dotp>0.9 and equal retention time to isotope labelled reference peptide.

FIGURE LEGENDS

Figure 1. Experimental plan for optimizing mutant protein detection in cell models. Two different variant calling platforms, *CLCBio* and *Varscan*, were used to identify DNA variants from the A375 cell line. Shotgun RNA seq was then performed on both parental and p53-null A375 cell lines to create a mutant expressed reference database for both cell lines. This mutant RNA database forms the reference from which mutant proteins will be identified using fractionation of the mutant tryptic peptides followed by mass spectrometry.

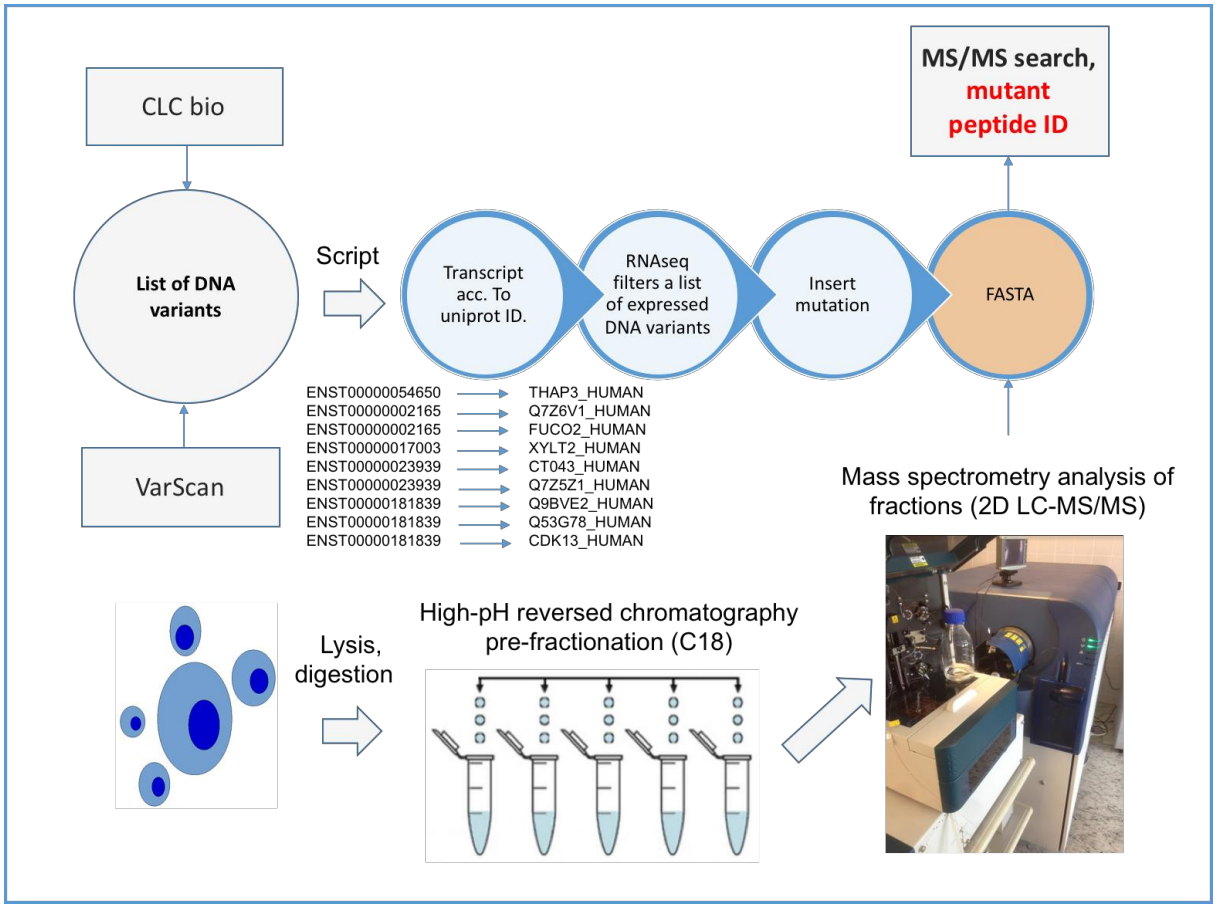


Figure 2. Representative variant calls from mutated genes. (A). Summary of the number of DNA mutations detected using the *CLCBio* and *Varscan* variant calling platforms. (B and C). Example *CLCbio* browser highlights mutated genes including a single nucleotide variant (MEIS1, chr2, 66691338) and an and in-frame insertion (SYNGR1, chr22, 39777822). The data are plotted by (i) chromosome position including hg19 reference genome (top, numbering); (ii) intron-exon boundaries. The thick line and thin line represent the exon and intron, respectively. Reference sets include: the blue line, that highlights ensemble v74 gene locus; the green line that represents ensemble v74 mRNA; and the yellow line that represents ensemble v74, CDS; (iii) highlights some of the DNA sequencing reads in blue aligned to hg19 with the DNA mutation variant (in A and B) highlighted by a color change and arrow**.** (D). *An analysis of shared DNA v RNA mutations in A375 cells*. A comparison of the overlap of variants detected in RNAseq and DNAseq filtered through distinct processes. The variants identified by DNAseq were filtered based on the presence of at least 4 mutant sequencing reads and at least a 50% frequency. A total of 1,468 variants were detected in 887 genes. The variants identified in mutated RNA were filtered by requiring at least

40 mutant RNA sequencing reads and at least 1 mutation in the DNAseq. This generated a different list of 1,418 mutated genes with highly expressed mutated RNA. Fusing the datasets produces a relatively small overlap of 107 common variants. The data suggest that the majority of mutated genes with a high confidence variant calling (774/887) are not highly expressed. *E. An example of one of the 107 shared mutated RNA and DNA CLCbio defined variants from the data filtered in Fig. 2D. GPATCH4 RNAseq* reads are highlighted as an example containing a frame-shift insertion mutation (AC) in DNA sequencing reads (Panel (I)) and in RNA sequencing reads (Panel (II)).

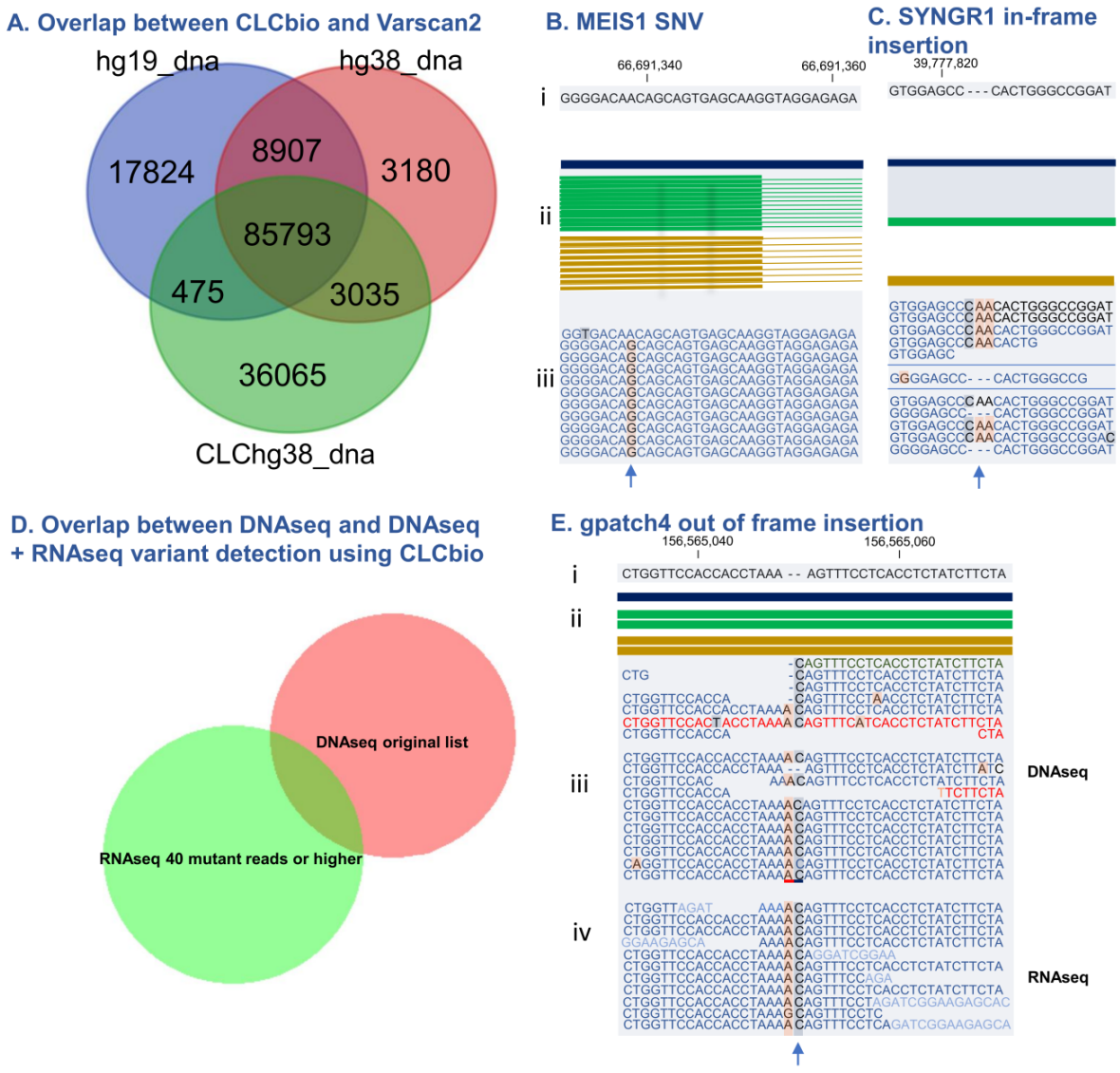
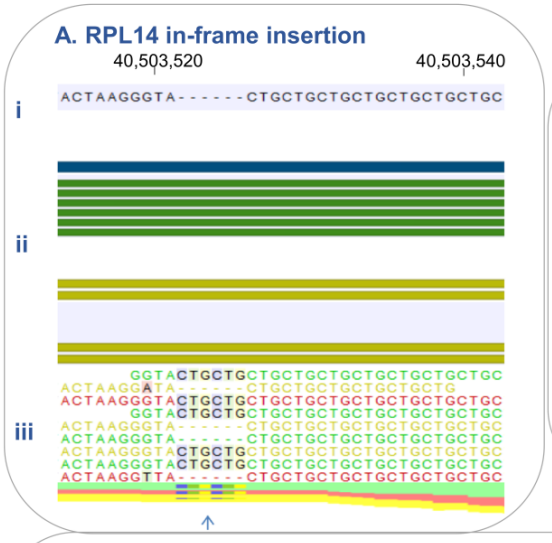


Figure 3. Identification of mutated rpl14 protein in A375 cells. The reference mutated DNA database was used to screen peptide spectra for evidence of mutant tryptic peptides. One mutant protein was detected by this method; rpl14. (A). The *rpl14* mutation is generated by an in-frame insertion of 6 bases encoding an AlaAla. The mutation is heterozygous as defined by the presence of 12 mutant reads and 11 wild-type reads (from Supplementary Table 1 and highlighted in the browser view by gaps in the sequence that matches the hg19 reference). The data are plotted by (i) chromosome position including hg19 reference genome (top, numbering); (ii) intron-exon boundaries. The thick line and thin line represent the exon and intron, respectively. Reference sets include: the blue line, that highlights ensemble v74 gene locus; the green line that represents ensemble v74 mRNA; and the yellow line that represents ensemble v74, CDS; (iii) highlights some of the DNA sequencing reads in blue aligned to hg19 with the DNA mutation variant highlighted by a color change and arrow. (B). The mutant protein tryptic coverage of rpl14 protein with the green highlighting high confident detection of two tryptic peptide sequences in (in green). (C and F). The mutant tryptic sequence containing 12 strings of alanine GTAAAAAAAAAAAK (mass 1156.60) has a 99% confidence of identification after fragmentation using ProteinPilot and there is no other such peptide in the human database. (D and E). The wild-type sequence containing 10 strings of alanine GTAAAAAAAAAAK (mass 957.5117) was also detected and has a 90% confidence of identification using ProteinPilot and there is no other such peptide in the human database. The existence of both tryptic peptides is consistent with the heterozygous nature of the tumour cell line (A).



B. Tryptic coverage of mutated RPL14 (ins AA)

MVFRRFVEVGRVAYVSFGPHAGKLVAIVDVIDQNRALVDP
CTQVRRQAMPFKCMQLTDFILKFPHSAHQKYVRQAWQK
ADINTKWAATRWAKKIEARERKAKMTDFDRFKVMKAKK
MRNRIIKNEVKKLQKAALLKASPKKAPGTKGTAAAAAAA
AAAAAAKVPAKKITAASKKAPAQKVPAQKATGQKAAPAP
KAQKGQKAPAQKAPAPKASGKKA

green = peptide confidence >95%
orange = peptide confidence <95% and >50%
red = peptide confidence < 50%
grey = no peptide coverage

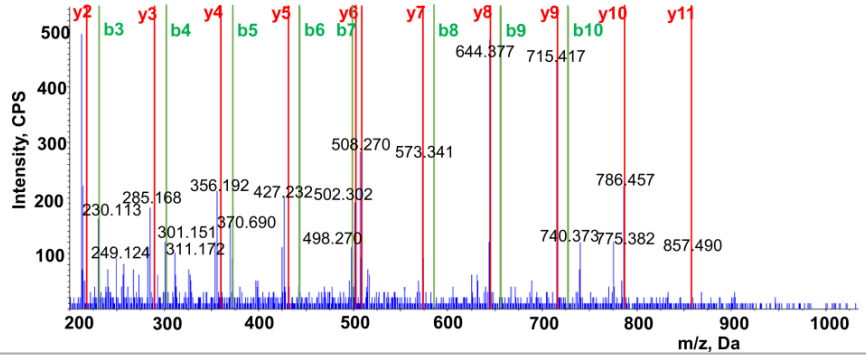
C. RPL14 (ins AA) GTAAAAAAAAAAAAK

Residue	Product Ion series b	b+2	y	y+2
1. G	58.0287	29.5180	1157.6273	579.3173
2. T	159.0764	80.0418	1100.6058	550.8066
3. A	230.1135	58.0287	999.5582	500.2827
4. A	301.1506	151.0790	928.5211	464.7642
5. A	372.1878	186.5975	857.4839	429.2456
6. A	443.2249	222.1161	786.4468	393.7271
7. A	514.2620	257.6346	715.4097	358.2085
8. A	585.2991	293.1532	644.3726	322.6899
9. A	656.3362	328.6717	573.7064	287.1714
10. A	727.3733	364.1903	502.2984	251.6528
11. A	798.4104	399.7089	431.2613	216.1343
12. A	869.4476	435.2274	360.2241	180.6157
13. A	940.4847	470.7460	289.1870	145.0972
14. A	1011.5218	506.2645	218.1499	109.5786
15. K	1139.6167	570.3120	147.1128	74.0600

D. RPL14 wild-type GTAAAAAAAAAAK

Residue	Product Ion series b	b+2	y	y+2
1. G	58.0287	29.5180	1015.5531	508.2802
2. T	159.0764	80.0418	958.5316	479.7694
3. A	230.1135	115.5604	857.4839	429.2456
4. A	301.1506	151.0790	786.4468	393.7271
5. A	372.1878	186.5975	715.4097	358.2058
6. A	443.2249	222.1161	644.3726	322.6899
7. A	514.2620	257.6346	573.3355	287.1714
8. A	585.2991	293.1532	502.2984	251.6528
9. A	656.3362	328.6717	431.2613	216.1343
10. A	727.3733	364.1903	360.2241	180.6157
11. A	798.4104	399.7089	289.1870	145.0972
12. A	869.4476	435.2274	218.1499	109.5786
13. K	997.5425	499.2749	147.1128	74.0600

E. Wild-type peptide GTAAAAAAAAAAK fragmentation spectrum



F. GTAAAAAAAAAAAAK (ins AA) fragmentation spectrum

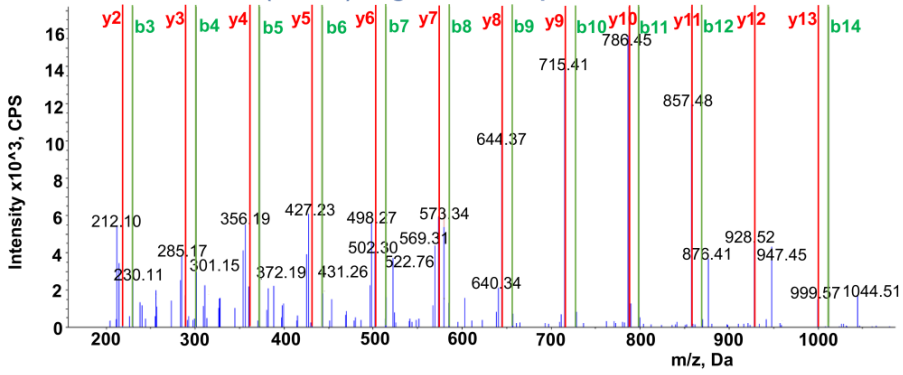


Figure 4. Optimization of mutant peptide identification using combined RNAseq and 2D LC-MS/MS datasets. A mutated reference search database file was based on RNAseq expression data (Supplementary Table 5) and then used to search 2D LC-MS/MS data leading to an increase in the number of peptides derived from potentially mutant proteins to 193 (Supplementary Table 9). The spectra in each of these were manually evaluated for the presence of y and b fragment ions that cover mutated amino acids characteristic by a m/z shift relative to product ions derived from the wild-type peptide. The data highlight successfully inspected mutated peptide candidates with an evidence of mutation in their peptide sequence. Tabulated are; Uniprot accession; mutation; mutated peptide sequence; peptide confidence; peptide charge state; peptide modifications and miscleavages; and product ions providing evidence of mutation (product ions that covered the mutant position in a peptide). The underlined peptides in the neoantigen sequence column highlights peptides that were subsequently validated using SRM.

Uniprot accession	Mutation	Mutated peptide sequence	Peptide confidence	Peptide charge state	Peptide modifications and misscleavages	Detected product ions proving evidence of mutation
H2A3	Val108Leu	LGRVTIAQGG**L**LPNIQAVLLPK	99	2+, 3+, 4+	cleaved L-L@N-term; missed R-V@3	b11, b12, b14, b15, b12 2+, b15 2+, b16 2+, b17 2+, b 19 2+, b20 2+, y11, y11 2+
SRP14	Pro124Ala	AAAAAAAAAPAAAATA**A**TTAATTAATAAQ	99	3+, 4+	-	y13, y14
CAN2	Asp22Glu	DREAAEGLGSHER	99	3+	missed R-E@2	y2, y3, y4, y5, y6, y7, y8
EZRI	Ile580Val	VDEFEA**L**	99	2+	-	b2, b3, b4, b5, b6
GSTO1	Ala140Asp	EDY**D**GLKEEFR	98	3+	missed K-E@7	b5, b7, y8, y8 2+, y9 2+, y10 2+
A0A0A0MTS2	Ile223Thr	TLAQLNPESSLFI**T**ASK	99	2+, 3+	-	y4, y5, y6, y7, y8, y9, y10, y11, y12, y13
TRAP1	Asp395Glu	LVSDGQALPEM**E**IHLQTNAEK	99	3+	-	y10, y11, y12, y13, y13 2+, y14 2+, y15 2+, y16 2+, y17 2+, y18 2+, y19 2+,
PHB	Arg43Leu	F**L**GVQDIVVGEGTHFLIPWVQKPIIFDCR	99	2+, 3+	Carbamidomethyl(C)@28	b2, b3, b4, b5, b6, b7
SPNXC	Leu68Val	TSPEEL**V**NDHAR	1	2+	missed R-E@12	y9
SUCB2	Thr396Ala	LEG**A**NVQEAQK	99	2+	-	b4, b5, b6, b7, b8, b9, b10, y8, y9, y10
TACC3	Ser190Phe	VSGSPEQAVEENLSSY**F**LDR	99	2+, 3+	-	y4, y5, y6, y7, y8, y9, y11
PLEC	Arg398Thr	SIITYVSSLYD**T**MPR	99	2+, 3+	-	y4, y5, y6, y7, y8, y9, y10, y11, y12, y13
MYH9	Ile1626Val	DLEAH**V**DSANK	99	2+, 3+	-	b6, b7, b8, b9, b10, y6, y7, y8, y9
RAD18	Arg302Gln	SAAEIV**Q**EIENIEK	99	2+, 3+	-	b8, b10, b12, b13, y8, y9, y10, y11, y12
PUR9	Thr116Ser	TVASPGV**S**VEEAVEQIDIGGVTLLR	99	3+, 4+	-	b8, b10, b11, b12, b13, b14, b15
B4DUC8	Val73Ile	NVDC**I**LLAR	99	2+	Carbamidomethyl(C)@4	b5, b6, b7, b8, y5, y6, y7, y8
ECHM	Thr75Ile	**I**FEEDPAVGAIVLTGGDK	99	2+, 3+	-	b2, b3, b4, b5, b6, b7, b8
PSD13	Asn13Ser	DVPGFLQQSQ**S**SGPGQPAVWHR	99	3+	-	y12, y13, y14, y12 2+, y14 2+, y15 2+, y16 2+, y17 2+, y18 2+, y19 2+, y20 2+
PDLI5	Ser492Asn	ILGEVI**N**ALK	99	2+	-	y4, y5, y6, y7, y8, y9, b8, b9
C1TC	Lys134Arg	DVDGLTSINAG**R**	99	2+	-	y4, y5, y6, y7, y8, y9, y10
K2C8	Arg369His	ASLEAAIADAEQ**H**GELAIK	99	2+, 3+	-	y7, y8, y9, y10, y11, y12, y13, y14, y15, y10 2+, y11 2+, y12 2+, y13 2+, y14 2+, y15 2+, y16 2+, y17 2+, y18 2+
GEMI4	Ala579Gly	FL**G**QILTAFPALR	99	2+, 3+	-	b3, b4, b5, b6, b7, b8, b9, y11, y12, y13 2+
PSB4	Ile234Thr	FQ**T**ATVTEK	99	2+	-	b3, b5, b6, b7, b9 2+, y7, y8
ANM3	Ser508Asn	VTVHK**N**K	84.8	2+	missed K-N@5	b6, y3, y4, y5, y6
A0A0A0MS30	Ala908Thr	LNGQQ**T**ALASQYR	98.8	2+	-	b10, y8, y9, y10, y13 2+
ESTD	Gly190Glu	AFS**E**YLGTDQSK	99	2+	-	b4, b5, b6, b7, b8, y9, y10, y12 2+
EMC1	Ser344Thr	**T**SSSEDGSMGSFSEK	99	2+	-	b3, b4, b5, b6, b7, b10 2+, b15 2+
AIP	Gln228Lys	EQPGSPEWIQLD**K**Q	98.2	2+	-	y2, y3, y4, y5, y6, y7, y8, y9, y10, y11, y12
RT27	Gly298Asp	EALDVL**D**AVLK	99	2+	-	b11 2+, y5, y6, y7, y8, y9, y10, y11 2+
UBQL4	Ile495Met	AMQALLQIQQGLQTLQTEAPGLVPSLGSFG**M**SR	95.5	4+	-	y4, y7, y8, y9, y10, y13, y14
SHCBP	Met21Thr	**T**GWAVEQELASLEK	99	2+	-	b3, b4, b5 2+
-	Ser259Gly	LQPLLNHLSHSYT**G**QDYSTQGNVGK	99	4+	-	b15 2+, b16 2+, b17 2+, b18 2+, y12, y14
RPC10	Ser24Ala	F**A**CNTCPYVHNITR	99	3+	Carbamidomethyl(C)@3, @6	b2, b3, b5, b12 2+
Q5SRN5	Glu176Val	WEAAH**V**AEQLR	64.1	3+	-	y6, y7, y8, y6 2+, y7 2+, y8 2+, y9 2+, y10 2+
CND1	Gln83Glu	SIDPGLKEDTL**E**FLIK	32	3+	missed K-E@7	b13 2+, b 15 2+, y5, y7, y8, y9, y10, y13 2+, y14 2+
PYGB	Lys622Asn	LII**N**LVTSIGDVVNHDPVVGDR	99	3+	-	b4
RD23B	Ala249Val	ESQAVVDPPQAASTG**V**PQSSAVAAAAATTTATT TTTSSGGHPLEFLR	99	4+, 5+	-	b16, b17, b33 2+, y33 2+
J3KQ32	Tyr274Cys	**C**DPGALVIPF	79.2	2+	Carbamidomethyl(C)@1, cleaved F-S@C-term	b2, b5, b6, b7, b8,
PPID	Leu302Ile	MSNWQGAIDSCLEALE**I**DPSNTK	99	3+	Carbamidomethyl(C)@11	y7, y8, y9, y10, y9 2+, y10 2+, y11 2+, y12 2+, y13 2+, y14 2+, y15 2+,
EFGM	Val215Ile	GI**I**DLIEER	99	2+	-	b3, b4, b5, b6, b7, y7, y7 2+, y9 2+
Q5SRN5	Asp185Glu	AYL**E**GTCVEWLR	99	2+	Carbamidomethyl(C)@7	b4, b5, b9, b11, b12 2+, y9, y10, y11
GRWD1	Arg8Pro	ESALEPG**P**VPEAPAGGPVHAVTVVTLLEK	98.5	3+	-	b9, b12, y22 2+, y24 2+, y25 2+
HEXA	Ile436Val	DFY**V**VEPLAFEGTPEQK	99	2+	-	b4, b5, b6, b4 2+
HEAT1	Asn1694Ser	NFGAENPDPFVPVL**S**TAVK	99	3+		y5, y6, y7, y8, y9, y10, y11, y7 2+, y8 2+
NEST	Val130Ala	AWLSSQ**A**AELER	90.2	2+	-	y6, y7, y8, y9, y10, b8
ADT2	Leu111Arg	YFAGNLASGGAAGATSLCFVYPLDFAR	99	3+	Carbamidomethyl(C)@18	b2, b3, b4, b5, b6, b7, b8, b10, b11, b12, b13, b14, b15, b16, b17, b9 2+, b10 2+, b11 2+, b13 2+, b14 2+, b15 2+, b17 2+, b18 2+, b20 2+
COPE	Thr117Ile	SVDVTN**I**TFLLMAASIYLHDQNPDAALR	99	3+, 4+	-	b10, b7 2+
EXOS7	Val274Leu	VLHASLQSV**L**HK	99	2+	-	b10, y7, y8, y9, y10, y10 2+, y12 2+
BIRC6	Val1332Leu	CAM**L**QFSEFHEK	1	2+, 3+	Carbamidomethyl(C)@1; Deamidated(Q)@5	b4, b5, y9, y12 2+
UBP24	Val2468Ala	NTFQLLHEILVIEDPIQ**A**ER	99	3+	-	y4, y6, y7, y8, y9, y10, y11, y6 2+, y10 2+
LGUL	Glu111Ala	ATLELTHNWGTEDD**A**TQSY	1	2+, 3+	cleaved Y-H@C-term	y19 2+
BRE1B	Gln615Arg	EREGPSLGPPPVASALSRADR	1	3+	missed R-E@2	b3, b5, b6, b7
SAM50	Ile345Val	FYLGGPTS**V**R	91	2+	-	y2, y3, y4, y5, y6, y7, y8, y9
PSB3	Met34Leu	FGIQAQ**L**VTTDFQK	99	2+	-	b7, b8, b11 b13, y9, y10, y11, y12, y13, y14 2+, y11 2+, y12 2+, y14 2+
C9JJ19	Leu109Phe	VRPDYTAQN**F**DHGK	99	3+	-	y5, y6, y7, y8, y9
IN35	Met128Val	VQVQPLELPMVTTIQVM**V**SSQLSGR	99	3+	-	y8, y9, y10, y11, y12, y13, y14
UBP24	Thr226Ile	NTFQLLHEILVIEDPIQ**A**ER	99	3+	-	y4, y6, y7, y8, y9, y10, y11, y6 2+, y10 2+
LAMC1	Leu888Pro	ACNCN**P**YGTMK	99	2+	Carbamidomethyl(C)@2; @4	y6, y7, y8, y9, y10 2+, y11 2+
RPL14	Ala157Ala-Ala ins	GTAAAAAAAAAA**AA**K	99	2+	-	b13, b14, y12, y13
HEAT1	His348Arg	LCNVPDLITILHGISETYDVSPLL**R**	99	3+	Carbamidomethyl(C)@2, cleaved H-L@N-term	y4, y6, y7, y11, y12, y13, y14

Figure 5. SRM validation of mutant peptide levels. *(A).* A List of identified mutated proteins and *(B).* corresponding peptides selected for validation using pseudo-selected reaction monitoring (pseudo-SRM). *(B).*highlights peptide sequences with green letters representing amino acids that are isotope labelled in the reference peptides. Red letters represent positions in peptide sequences that are mutated. *(C).* and *(D).* represent optimization of isotope labelled reference peptide QRVDEFEAL spiked-in into the A375 lysate. Extracted product ion chromatogram *(C).* shows intensity of SRM transitions from QRVDEFEAL mutated reference isotope labelled peptide at various concentrations, while D shows peak areas of corresponding product ions (y3 - y8). From *(C).* and *(D).* we determined that optimal spike in of QRVDEFEAL reference peptide is approximately 14.3 fmol which corresponds to a 10,000x dilution of stock reference peptide solution. This same procedure was carried out for all 10 heavy labelled peptides spiked into A375 lysates.

A. Mutated peptide chosen for p-SRM

Protein	Neopeptide sequence	Mass	Charge	Mutation
Keratin, type II cytoskeletal 8_KRT8	ASLEAAIADAEQHGELAIK	1935,9811	3+,2+	R to H
HLA class I histocompatibility antigen_HLA-A	WEAAHVAEQLR	1495,7169	2+	D to E
ADP/ATP translocase 2_SLC25A5	YFAGNLASGGAAGATSLCFVYPLDFAR	2795,3381	3+	L to R
Ezrin_EZR	QRVDEFEAL	1088,5101	2+, 2+	I to V
Ezrin_EZR	VDEFEAL	821,3779	2	I to V
Plectin_PLEC	SIITYVSSLYDTMPR	1744,8699	2+,3+	A to T2
Glycogen phosphorylase, brain form_PYGB	LIINLVTSIGDVVNHDPVVGDR	2344,2761	3+	K to N1
RPL14 protein_RPL14	GTAAAAAAAAAAAAK	1156,6073	2+	AA insertion
Transforming acidic coiled-coil-containing protein 3_TACC3	VSGSPEQAVEENLSSYFLDR	2226,0449	3+	S to F
PDZ and LIM domain protein 5_PDLIM5	ILGEVINALK	1068,6533	2+	S to N

B. Heavy mutated peptide chosen for p-SRM

Peptide Name	Sequence	Peptide Length	Stable Isotope-Labeled Residue
1	ASLEAAIADAEQHGELAIK	19	Lysine (K), +8Da
2	*WEAAHVAEQLR	11	Arginine (R), +10Da
3	*YFAGNLASGGAAGATSLCFVYPLDFAR	27	Arginine (R), +10Da
4	QRVDEFEAL	9	Arginine (R), +10Da
5	VDEFEAL	7	Leucine (L), +7Da
6	SIITYVSSLYDTMPR	15	Arginine (R), +10Da
7	LIINLVTSIGDVVNHDPVVGDR	22	Arginine (R), +10Da
8	GTAAAAAAAAAAAAK	15	Lysine (K), +8Da
9	VSGSPEQAVEENLSSYFLDR	20	Arginine (R), +10Da
10	ILGEVINALK	10	Lysine (K), +8Da

* new cleavage site

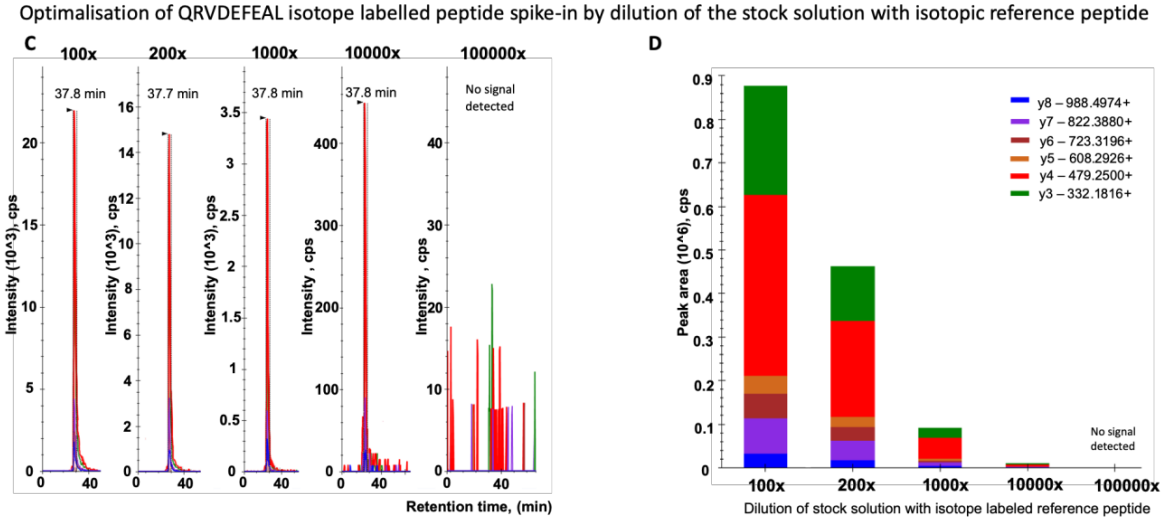


Figure 6. Extracted product ion chromatograms for SRM validated mutated peptides. *(A).* and *(B).* a list of product ions representing two selected high confidence mutated peptides (peptide confidence > 99%) identified using the 2D LC-MS/MS data. Product ions labelled with brown color were identified in the LC-MS/MS spectra of VSGSPEQAVEENLSSYFLDR and IIPTVLMTEDIK peptides (A. and B., respectively). Underlined product ion masses represent identified product ions that confirm presence of the amino acid mutation in the peptide sequence. *(A).* Fragmentation evidence of the VSGSPEQAVEENLSSYFLDR mutated peptide highlights 11 product ions confirming the S to F mutation. *(B).* Fragmentation evidence of the IIPTVLMTEDIK mutated peptide identifies only 3 product ions confirming the A to P mutation. The fragmentation evidences clearly highlight that the probability of amino acid mutation is not reflected in peptide confidence; therefore, these two peptides are shown as an example that it is important to evaluate the spectra/fragmentation evidence to determine mutant sequence status. *(C).* and *(D).* example of pseudo-SRM quantitation of the SIITYVSSLYDTMPR peptide referencing the heavy and light peptide titrations and defining key features required for high confidence validation (1-5) and other peak features (1-3). Left section of *(C).* highlighted in red shows product ions representing light and heavy form of peptide. While product ion chromatograms show MS/MS signal of these product ions across the LC run. *(D).* Tabulated results from a pseudo-SRM validation of selected mutated peptides using isotope labelled reference peptides. The table highlights sequences and charge states of the mutated peptides (the first and the second columns) followed by columns with key characteristics of the product ions representing intrinsic or reference mutated peptides in a sample. The third column defines the consistency of retention times among the reference and intrinsically mutated peptides. The intrinsic peptide must have an identical retention time to the reference peptide as its chromatographic characteristics remain unchanged by isotope labelling. The fourth column defines the intensity of the intrinsic and the reference mutated peptide. Both peptides must have a sufficient intensity (peak height) to be considered as detected. A peptide is defined as detected when its intensity is at least 3 times the background noise of the method (noise of pseudo-SRM was approximately 100 counts per second) according to FDA directions. The fifth column filters out wild-type peptides; the spectra of which could overlay with the spectra of the mutated peptides. We list detected product ions as having the characteristic mass shift of a mutated amino acid; this further validates the existence of a mutated peptide. The higher the number of product ions with the characteristic mass shift of mutated amino acid, the higher the confidence of a mutated peptide. The sixth column describes the similarity of an intensity pattern between the intrinsic and the mutated peptide. The ionization efficiency of both peptides must be the same, therefore we expect the same intensity rank of product ions in both peptides. We inspected corresponding product ion chromatograms and we selected only product ions that show similar patterns in terms of the intensity. The higher the number of product ions the higher the similarity between the intrinsic and the reference mutated peptide. The seventh column shows dotP value. This value reflects and summarizes the similarity between the intrinsic and the reference mutated peptide. It ranges in between 0 to 1 and the higher the value is, the higher the similarity of both peptides. Usually, pseudo-SRM validated peptides have high dotP values. The last column summarizes values determined from the pseudo-SRM assay. Peptides with an excellent pseudo-SRM evidence could be considered as certainly validated. **Abbreviations:** L = (light) naturally occurring intrinsic form of mutated peptide, H = (heavy) reference isotope labelled peptide, WT = wild-type, RT = retention time.

A

TACC3 (S to F) VSGSPEQAVEENLSSYFLDR

Residue	Product Ion series b	b+2	y	y+2
1. V	100.0757	50.5415	2227.0513	1114.0293
2. S	187.1077	94.0575	2127.9829	1064.4951
3. G	244.1292	122.5682	2040.9509	1020.9791
4. S	331.1612	166.0842	1983.9294	992.4684
5. P	428.2140	214.6106	1896.9874	948.9523
6. E	557.2566	279.1319	1799.8446	900.4260
7. Q	685.3151	343.1612	1670.8020	835.9047
8. A	756.3523	378.6798	1542.7435	771.8754
9. V	855.4207	428.2140	1471.7064	736.3568
10. E	984.4633	492.7353	1372.6379	686.8226
11. E	1113.5059	557.2566	1234.5953	622.3013
12. N	1227.5438	614.2780	1114.5527	557.7800
13. L	1340.6329	670.8201	1000.5098	500.7585
14. S	1427.6649	714.3361	887.4258	444.2165
15. S	1514.6969	757.8521	800.3937	400.7005
16. Y	1677.7602	839.3838	713.3617	357.1845
17. F	1824.8286	912.9180	550.2984	275.6528
18. L	1937.9127	969.4600	403.2300	202.1186
19. D	2025.9397	1026.9735	290.1459	145.5766
20. R	2209.0408	1105.0240	175.1190	88.0631

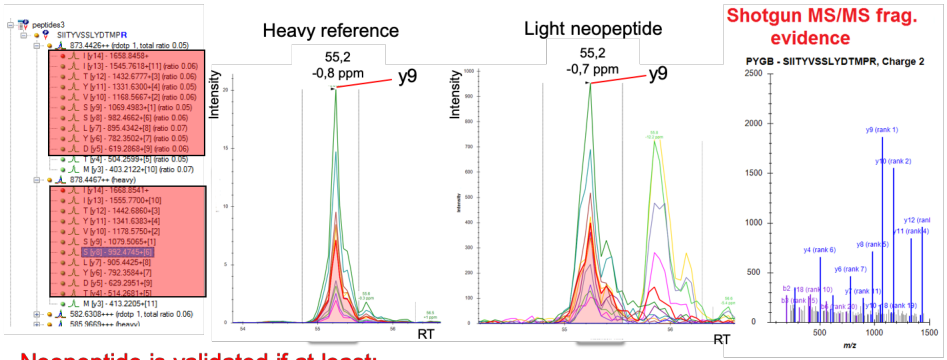
B

EIF4G1 (A to P) IIPTVLMTEDIK

Residue	Product Ion series b	b2+	y	y2+
1. I	114.0913	57.5493	1372.7756	686.8914
2. I	227.1754	1140913	1259.6916	630.3494
3. P	324.2282	162.6177	1146.6075	573.8074
4. T	425.2758	213.1416	1049.5547	525.2810
5. V	524.3443	262.6758	948.5070	474.7572
6. L	637.4283	319.2178	849.4386	425.2230
7. M	768.4688	384.7380	736.3546	368.6809
8. T	869.5165	435.2619	605.3141	303.1607
9. E	998.5591	499.7832	504.2664	252.6368
10. D	1113.5360	557.2967	375.2238	188.1155
11. I	1226.6701	613.8387	260.1969	130.6021
12. K	1354.7551	677.8862	147.1128	74.0600

C

SIITYVSSLYDTMPR (Fraction - 65% B)



Neopeptide is validated if at least:

1. same fraction
2. equal RT
3. Similar fragmentation pattern
4. Similar fragment ion intensities
5. Same m/z shift

and

Other peak-group features:

1. Coelution
2. Shape
3. Summed intensity

D

Mutated peptide candidate identified in 2D LC-MS/MS screen	Precursor ion charge state	RT consistency among H and L mutated peptide (min)	Peak height of the most intensive product ion A375 wt/ A375 p53 null (cps)	Product ions proving the evidence of mutated aminoacid in peptide	Product ions with similar intensity patterns in H and L mutated peptide	dotP value	Overall comment on mutated peptide p-SRM evidence
ASLEAAIADAEQ**H**GELAIK	2+, 3+	yes (46)	1500/1000	7	y12, y11, y7, y6, y10	1	excellent
WEAAH**V**AEQLR	3+	yes (33)	1000/2000	6	y5, y6, y4, y3, y7, y8	1	excellent
YFAGNLASGGAAGATSLCFVYP LDFAR (new trypsin cleavage site before Y)	3+	yes (55)	7000/9000	14	y7, y6	0.89	fair
QRVDEFEA**L**	2+	yes (39)	2500/2500	6	b7, b8, b5, b6, b4, b3	1	excellent
VDEFEA**L**	2+	yes (43)	500/500	4	b6, b5, b3, b4, y3, y4	1	excellent
SIITYVSSLYD**T**MPR	2+, 3+	yes (55)	3000/3000	11	y9, y10, y12, y11	1	excellent
LII**N**LVTSIGDVVNHDPVVGDR	2+, 3+	yes (54)	500/1400	4	y6, y13, y9, y10,	1	excellent
GTAAAAAAAAA**A**K	2+, 3+	yes (32)	7000/3500	2	y10, y9, y11, y8, y7, y6,	1	excellent
VSGSPEQAVEENLSSY**F**LDR	3+	yes (54)	100/200	12	y7, y6	0.89	weak
ILGEVI**N**ALK	2+	yes (44)	2000/2000	6	y9, y8, y6, y5, y4, y7	1	excellent

Figure 7. Generating a melanoma cell line with a p53-null status using CRISP-R mediated gene editing. *(A). Position of guide RNA-binding motif.* Guide RNAs targeting exon 5 of the p53 gene were cloned into pBT-U6-CAS9-2A-GFP. The intron-exon-intron sequence of the p53 gene is highlighted along with the position and orientation of the guide RNA (reverse arrow; in blue with the PAM sequence in red). A375 melanoma cells were transfected with the guide RNA and clones were selected using GFP selection by FACS followed by plating for single cells. Individual cells were grown and screened for evidence of gene editing and p53 activation status. *(B and D). DNA sequencing*. Sequence of the chromosomal DNA derived from the p53 knock-out cells by amplification of the chromosomal region using PCR primers flanking the guide RNA target motif. *(C and E). Translation of the gene edits produces theoretical stop codons*. *(F and G). Immunoblotting to define the p53 and HLA expression status.* (F). Lysates from A375 (wt and p53-null) and HCT116 cells (wt and p53-null) were immunoblotted with the p53 antibody DO-1. This antibody binds to the N-terminal domain of p53. A loading control was detecting PCNA using the PC10 antibody. *(G).* Cell lysates from the indicated A375 variants (wt and -/-), as SiHa cells (as a control, C) were immunoblotted with antibodies to MHC Class I allele, HLA-B. Actin was used as a loading control.

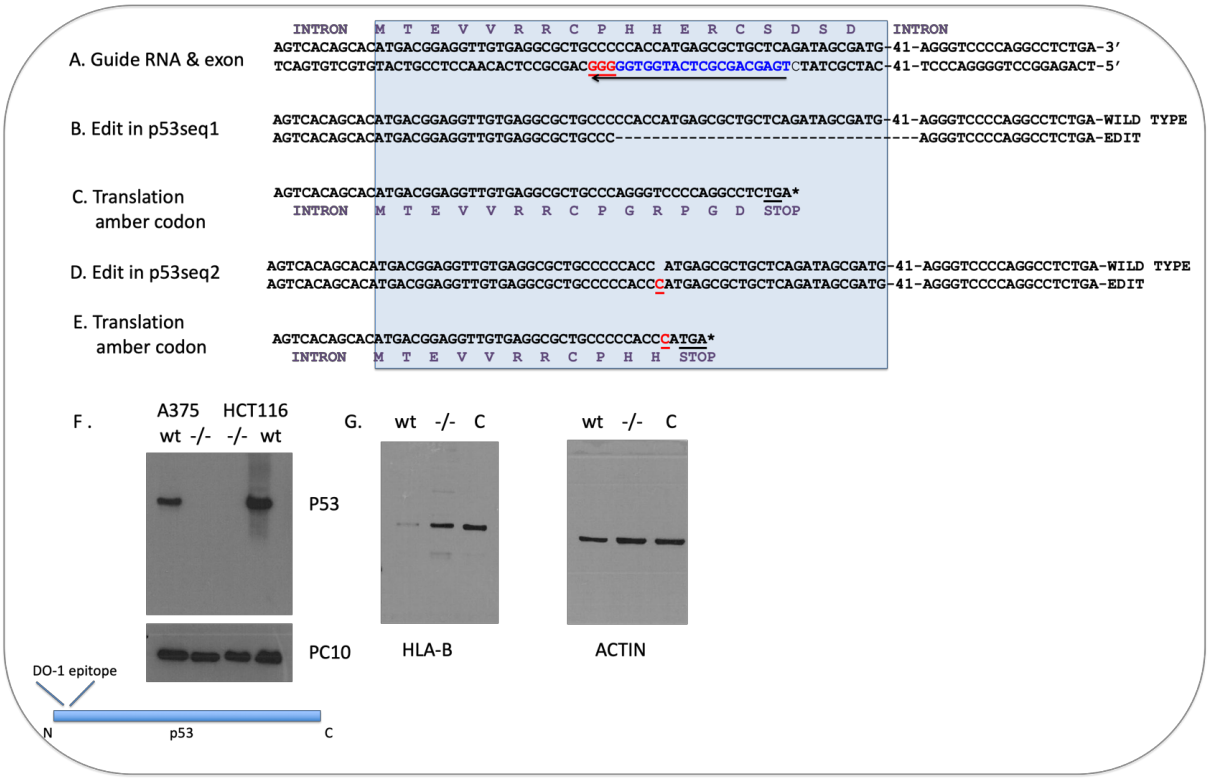


Figure 8. SWATH quantitation pipeline of mutated peptides in isogenic p53 wt and null cell panels. *(A).* An example of extracted product ion chromatogram corresponding to a mutated peptide from Glycogen phosphorylase, brain form (PYGB) detected using SWATH acquisition in A375 cell lysate. A legend in the right shows rank of the most intensive product ions from mutated peptide. *(B).* A spectral library MS/MS spectrum corresponding to peptide LIINLVTSIGDVVNHDPVVGDR. An intensity-based product ion rank in spectral library should be identical to product ion rank in product ion chromatogram to consider peptide hit as valid. *(C).* Quantification using three technical replicates between A375 p53-null cells and A375 p53 wild-type cells shows an up-regulation of mutated peptide in A375 null cells. Here, as problematic we see a quantification on a set of product ions which might refer also to wild-type peptide form. Therefore, we recommend inspecting a set of selected product ions for quantitation and filter out only product ions uniquely covering the mutant position in a peptide sequence. In case of PYGB LIINLVTSIGDVVNHDPVVGDR mutated peptide it is b4 – b21 and y19 – y21 and in case of PLEC mutated peptide SIITYVSSLYDTMPR it is b12 – b15 and y4 – y15. *(D).* Extraction of product ion chromatogram for SIITYVSSLYDTMPR mutated peptide (PLEC). SIITYVSSLYDTMPR quantitation refers more uniquely to the mutated peptide form compared to quantitation of LIINLVTSIGDVVNHDPVVGDR peptide. It relies on product ions that are characteristic exclusively for mutant peptide form (y4, y5, y6, y7). *(E).* Shows corresponding spectral library for SIITYVSSLYDTMPR mutated peptide and intensity rank of product ions. *(F).* Shows quantitation of SIITYVSSLYDTMPR mutated peptide in three technical replicates of A375 p53-null cells and A375 p53 wild-type cells. Relative quantitation of both mutated peptides in A375 p53- and A375 WT cell line *(C).* and *(F).* was determined from a sum of 4 product ion peak areas by integration of AUC. Bar graph *(C* and *F).* represents peak areas corresponding to these 4 product ions. Overall peptide quantity in each replicate is represented by an entire bar. *(G).* SWATH quantitation of a control peptide from ACTB to evaluate the effects of sample preparation and differences between mass spectrometry measurements.

Mutated peptide quantitation using SWATH acquisition in A375 cell lines

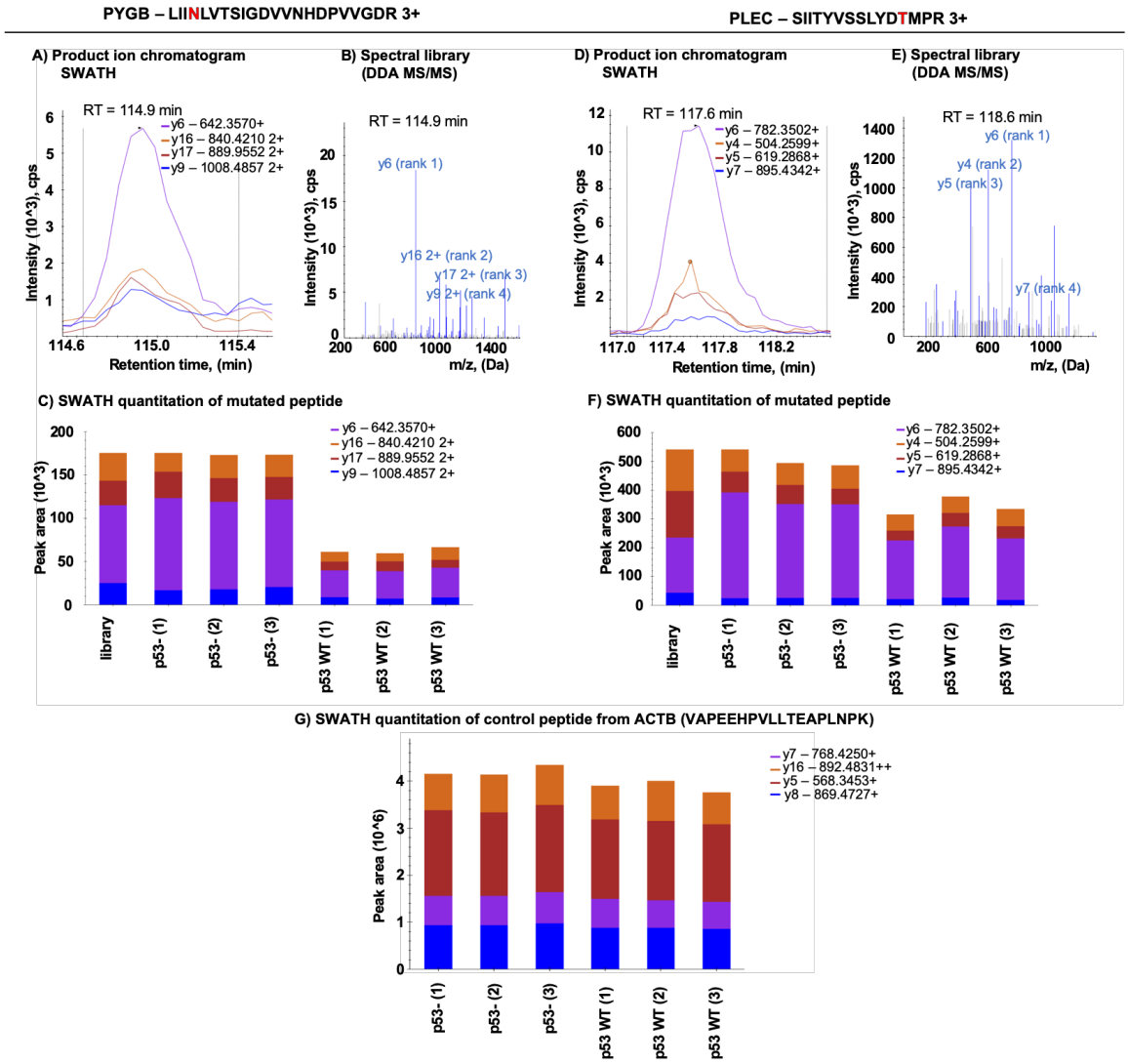


Figure 9. Summary of mutated peptides detected using the optimized platforms. Using both CLCbio (CLC) and Varscan (VS), variants were detected requiring at least 10 RNA mutant reads and 1 DNA mutant read. A. A comparison of the enriched mutant peptides detected in wt-p53 vs p53-null cells using mass spectral data summarized in Supplementary Tables 13 and 14. B and C. Summarizes the mutant tryptic peptides detected in wt-p53 or p53-null cells, using CLCbio and Varscan platforms. The reference database generated from the DNA and RNAseq was converted to protein amino acid sequences by TransPEM [47] and subsequently this reference database was used in ProteinPilot 4.5 to search for spectra with matches that define mutated peptides.

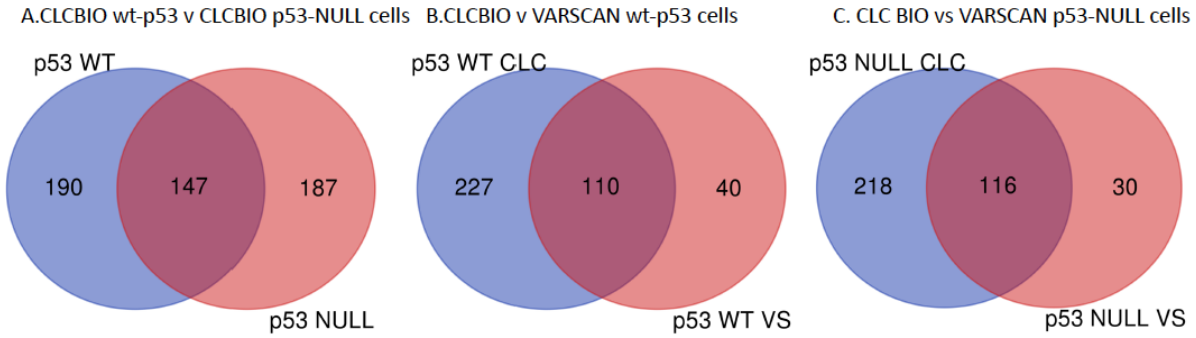


Figure 10. STRING analysis of mutant protein networks in wt-p53 cells. STRING (https://string-db.org) is a database of protein-protein interactions including direct experimental data, indirect functional associations, from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from additional primary databases [48]. The known interactions are defined as: curated databases (——), and experimentally determined (——). The predicted interactions are defined as: gene neighborhoods (——), gene fusions (——), gene co-occurance (——), textmining (——), co-expression (——), and protein homology (——). The minimum required interaction score was defined as high confidence (0.7). STRING was used to define interaction networks composed of mutant proteins in the wt-p53 cell tryptic peptide dataset that differ from p53-null cells. The mutant peptides from wt-p53 cells are summarized in Supplementary Table 13.

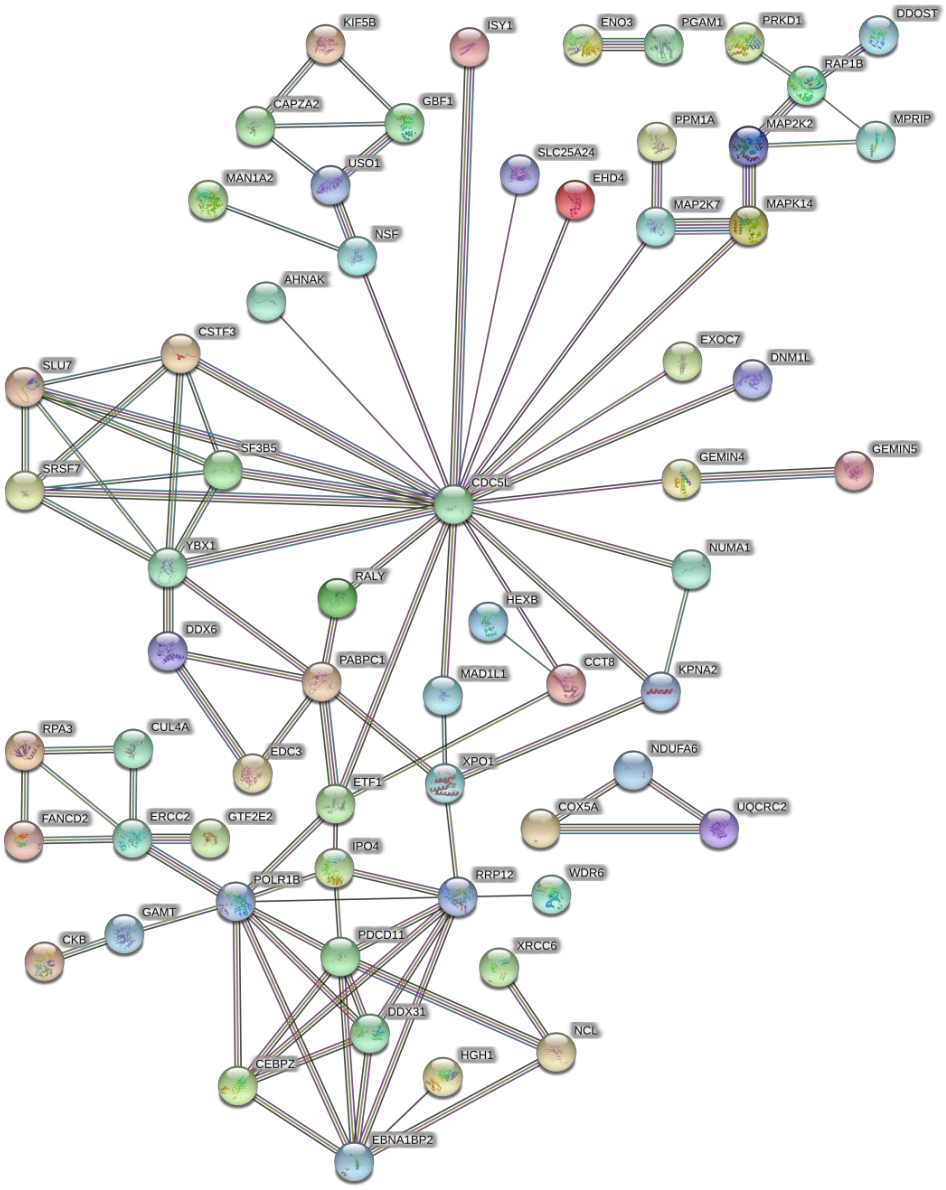


Figure 11. STRING analysis of mutant protein networks in p53-null cells. STRING was used as in Fig. 10 to define interaction networks composed of mutant proteins in the p53-null cell tryptic peptide dataset that differs from wt-p53 cells. The known interactions are defined as: curated databases (——), and experimentally determined (——). The predicted interactions are defined as: gene neighborhoods (——), gene fusions (——), gene co-occurance (——), textmining (——), co-expression (——), and protein homology (——). The minimum required interaction score was defined as high confidence (0.7). The mutant peptides from p53-null cells are summarized in Supplementary Table 14.

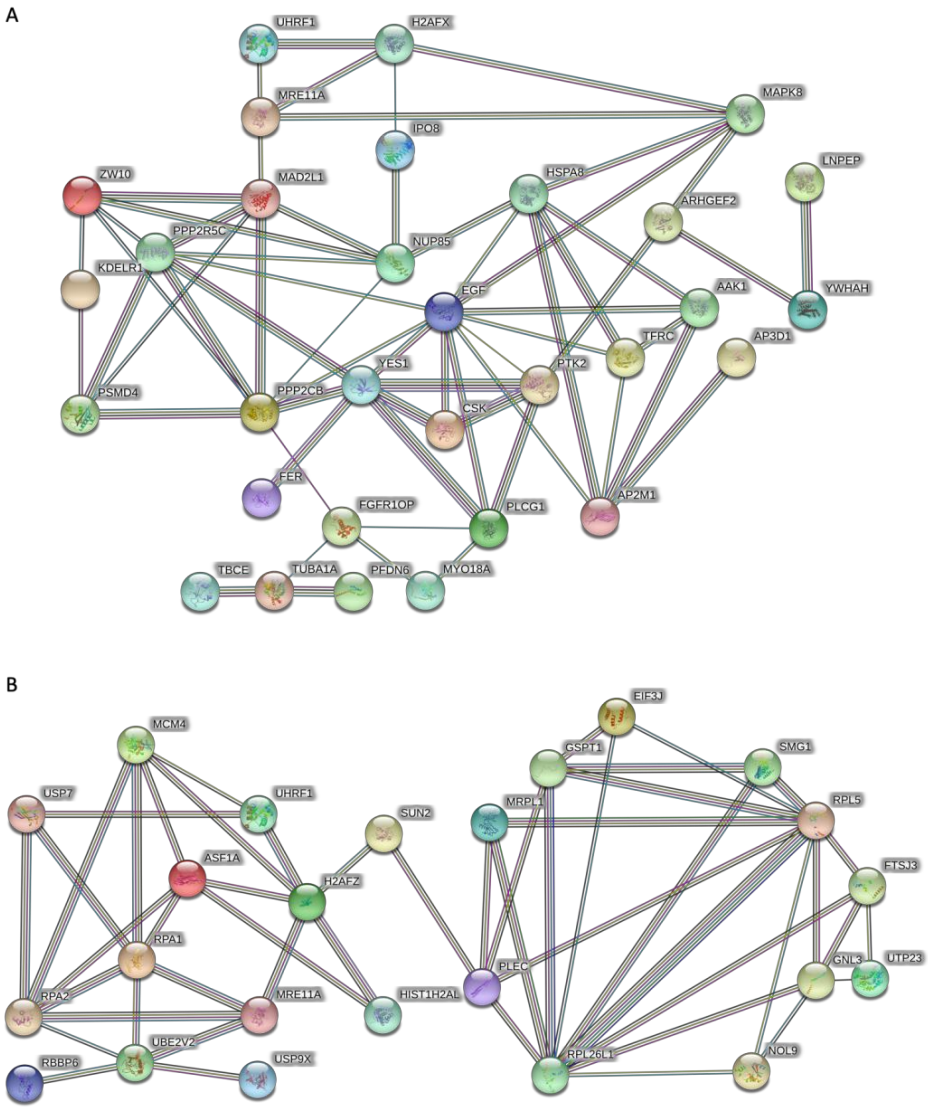


TABLE 1. SWATH-MS analysis of proteins whose genes are mutated using wt-p53 and p53-null A375 cells. The table highlights gene name, amino acid substitution, and fold change in the expression of wt vs p53-null cells. Only one protein on this list has a mutated peptide detected using mass spectrometry (rpl14, Fig. 3). The other 18 proteins with homozygous mutations do not have any tryptic mutated peptides detected using SWATH-MS.

N	Gene name	Amino acid change(s)	Fold Change	N	Gene name	Amino acid change(s)	Fold Change
1	**SEC31A**	Asp22Tyr	**2.74**	22	**PRKCSH**	Pro33Leu	**0.98**
2	**PYGB**	Lys622Asn	**2.42**	23	**TCP1**	Lys321Arg	**0.98**
3	**ADPGK**	Ala8Val	**2.34**	24	**SPTBN1**	Pro883Thr	**0.98**
4	**EGFR**	Arg521Lys	**2.29**	25	**PABPC1**	Leu593Val Leu597Pro Met584Ile Leu562Ser Val517Leu Arg506Cys Val505Ile Arg493Cys Pro402Leu Arg374Cys Glu372Gly Met251Ile Ala154Gly Thr147Met His144Arg Cys132Gly	**0.97**
5	**LYAR**	His265Arg (275 in this transcript)	**2.11**	26	**PSMA2**	Glu3Lys	**0.97**
6	**NDUFA10**	Phe169Leu	**1.59**	27	**PDHA1**	Arg27fs	**0.96**
7	**HADH**	Leu90Pro	**1.52**	28	**HSP90AB1**	Arg719His	**0.93**
8	**LAMC2**	Asp784Asn	**1.33**	29	**MARS**	Thr32Ser	**0.93**
9	**LRRC59**	Glu69Asp Lys137_Pro138insGln	**1.29**	30	**PPP1CA**	Phe225fs	**0.92**
10	**SLC25A5**	Gly73Ser Asn77Thr	**1.23**	31	**NCL**	Asp258del	**0.91**
11	**BCLAF1**	Thr888Asn	**1.19**	32	**ACTG1**	Ser185Cys Glu170Val	**0,90**
12	**WDR1**	Ile185Val	**1.16**	33	**PRSS1**	Asn29Ile Cys185Tyr Ser195Asn Met197Val	**0.89**
13	**SNRPC**	Cys13_Arg14insHis Arg14Leu	**1.14**	34	**RPL14**	Ala159_Lys160insAlaAla	**0.86**
14	**PDCD6IP**	Val378Ile	**1.13**	35	**AHCY**	Arg257Gln	**0.85**
15	**NIT2**	Phe27Ser	**1.10**	36	**SRRM2**	Arg581Trp	**0.82**
16	**PHB2**	Arg123Cys	**1.09**	37	**FUBP1**	Gly455Trp	**0.80**
17	**HEATR1**	**Val1854Ala**	**1.09**	38	**PPA2**	Lys282Asn	**0.72**
18	**RBMX**	**Asn1Lys**	**1.05**	39	**ARHGEF2**	Gly582Trp	**0.64**
19	**NACA**	**Pro649Ser**	**1.03**	40	**FAM162A**	Arg142Cys	**0.54**
20	**PFDN5**	**Pro65Ser**	**1.03**	41	**GC**	His445Arg	**0.40**
21	**PLEC**	**Ala398Thr**	**0,99**	42	**MPRIP**	Ser189del	**0.20**

SUPPLEMENTAL LEGENDS

Supplementary Table 1A. Excel file showing a list of mutated genes with mutations detected at a 5% frequency or higher in DNA from A375 cells using the *CLCbio* variant platform detector. The analysis identified 120,325 genes. Represented are data including chromosome number and position, reference allele and mutation type, zygosity, Count (mutation number) and coverage (total sequencing reads), frequency of mutation, gene cards name, ensemble qualifiers, amino acid change, coding region change (Syn or non-syn), and dbSNP identifiers. **Supplementary Table 1B**. Excel file showing a list of mutated genes detected at a 40% frequency or higher in DNA from A375 cells (from Supplementary Table 1A) using the *CLCbio* variant detector. This analysis generated 63,880 variants with mutations. The data are plotted as a function of; chromosome; region; type of mutation; reference base; allele base; reference allele; zygosity; count; coverage; frequency of mutation; gene card name; ensemble name; gene name; gene biotype; transcript name; coding region change; amino acid change; amino acid change in longest transcript; other variants in codon; whether the mutation is non-synonymous, synonymous, or out with exons; and dbsnp reference.

Supplementary Table 2. Excel file showing a list of mutated genes with non-synonymous mutations detected at a 40% frequency or higher in A375 cells (from Supplementary Table 1B) using the *CLCbio* variant detector. This analysis generated 1,468 genes with non-synonymous mutations. The data are plotted as a function of; chromosome; region; type of mutation; reference base; allele base; reference allele; length of change; zygosity; count; coverage; frequency of mutation; probability; forward read count; reverse read count; ratio; average quality; exact match to dbsnp; gene card name; ensembl name; gene name; gene biotype; transcript name; coding region change; amino acid change; amino acid change in longest transcript; other variants in codon; whether the mutation is non-synonymous, synonymous, or out with exons; and dbsnp reference.

Supplementary Table 3. Excel file showing RNA sequencing reads mapped to the human reference genome hg19, with dbSNPs removed, to identify mutated and expressed genes. A total of 18,341 expressed non-synonymous mRNA variants were identified with a cutoff of 5% of the total RNA reads defined as mutated. The data are plotted as a function of; chromosome; position; type of mutation; reference; allele; reference allele; length of change; zygosity; RNA count; RNA coverage; RNA frequency; RNA forward reads; RNA reverse reads; DNA read count; DNA read coverage; gene card name; ensemble name; gene name; transcript ID; coding region change; amino acid change; and whether the mutation is non-synonymous, synonymous, or out with exons.

Supplementary Table 4. Excel file showing RNAseq data with a stringent cutoff requiring at least 40 mutated RNA sequencing reads and filtered against genomic DNA cutoff of at least 1 mutant genomic DNA sequencing read. This produced 5,980 RNA variants including synonymous, non-synonymous, and non-exonic mutations. The data are plotted as a function of; chromosome number; type of mutation; zygosity; gene card name; ensemble name; gene name; amino acid change; and whether the mutation is non-synonymous, synonymous, or out with exons.

Supplementary Table 5. Excel file showing RNAseq data with a stringent cutoff requiring at least 40 mutated RNA sequencing reads and filtered against genomic DNA cutoff of at least 1 mutant genomic DNA sequencing read (from Supplementary Table 4). Upon filtering for non-synonymous variants, a list was generated composed of 1,418 non-synonymous highly expressed RNA variants. This provides a conservative estimate of the number of mutated expressed mRNA in the A375 cell

line, as there are many mutated mRNAs quantified at reads from 39 and lower. However, we focus on those mutant mRNA species which are abundant, perhaps not degraded by NMD or perhaps not resulting from expression of minor subclones in the A375 cell population.

Supplementary Table 6. Excel file showing a subset of genes acquired after the list composed of 1,418 non-synonymous highly expressed RNA variants (Supplementary Table 5) with the 1,468 CLC genomic DNA variants using the stringent DNA variant calling (Supplementary Table 2). The excel file highlights a total of 877 out of 985 genes with expressed mRNA variants are not present in the original DNA variant list.

Supplementary Table 7. Excel file showing the tryptic peptides processed using SWATH mass spectrometry to identify proteins differentially expressed in the wt and p53-null cells using the *normal* reference proteome Swiss-prot and TrEMBL (potential mutant proteins from this are listed in Table 1).

Supplementary Table 8. Excel file showing tryptic peptides identified from shotgun mass spectrometry including a pre-fractionation step to increase the coverage of total peptides to over 35,000 with a coverage of over 4,500 proteins. (A). represents data from wt-A375 cells and (B). represents data from p53-null A375 cells.

Supplementary Table 9. Excel file showing a subset of proteins listed by selecting proteins from the expressed (mRNA) and mutated genes detectable in the shotgun 2D LC-MS/MS experiment (Supplementary Table 8A and 8B) using the RNA variant file (Supplementary Table 5). Applying RNAseq derived mutant search database to search the 2D LC-MS/MS data, this increased the number of detectable mutant peptides to 193. Of this selection, we manually validated all spectra from the 193 peptides and we produced data (Fig. 4) of 60 mutated proteins of relatively high confidence. Of these 60 mutant proteins, 10 were validated by SRM (Fig. 5; Supplementary Table 10) to produce a conservative FDR of 20%. Thus, 4 out of 5 of the manually validated, high confidence mutated peptides are likely to be mutated.

Supplementary Table 10. Excel file showing results summary of SRM mass spectrometry analysis in a subset of 10 mutated proteins. Evaluating similarity of product ion peakgroups referring to neopeptides is summarised by dotp value ranging from 0 to 1. Neopeptides with dotp above 0.9 and with more than 2 product ions above LOD (three times noise) are considered as successfully validated. Abbreviation H refers to spiked-in "heavy" isotope labelled neopeptide and L refers to endogenous "Light" form of neopeptide.

Supplementary Table 11. A list of DNA variants detected using VarScan2 using a minimal coverage of 1 and hg38.
Supplementary Table 12. A list of RNA variants detected using VarScan2 using a minimal coverage of 10 and hg38.
Supplementary Table 13. Mutant tryptic peptide lists derived from lysates from wt-p53 cells using CLCbio software.
Supplementary Table 14. Mutant tryptic peptide lists derived from lysates from p53-null cells using CLCbio software.

REFERENCES

[1] G.R. Bignell, C.D. Greenman, H. Davies, A.P. Butler, S. Edkins, J.M. Andrews, G. Buck, L. Chen, D. Beare, C. Latimer, S. Widaa, J. Hinton, C. Fahey, B. Fu, S. Swamy, G.L. Dalgliesh, B.T. Teh, P. Deloukas, F. Yang, P.J. Campbell, P.A. Futreal, M.R. Stratton, Signatures of mutation and selection in the cancer genome, Nature, 463 (2010) 893-898.
[2] M.R. Stratton, Journeys into the genome of cancer cells, EMBO Mol Med, 5 (2013) 169-172.
[3] J.W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical R&D efficiency, Nat Rev Drug Discov, 11 (2012) 191-200.
[4] M. Paoloni, C. Khanna, Translation of new cancer treatments from pet dogs to humans, Nat Rev Cancer, 8 (2008) 147-156.
[5] L.B. Alexandrov, P.H. Jones, D.C. Wedge, J.E. Sale, P.J. Campbell, S. Nik-Zainal, M.R. Stratton, Clock-like mutational processes in human somatic cells, Nat Genet, 47 (2015) 1402-1407.
[6] S. Nik-Zainal, J.E. Kucab, S. Morganella, D. Glodzik, L.B. Alexandrov, V.M. Arlt, A. Weninger, M. Hollstein, M.R. Stratton, D.H. Phillips, The genome as a record of environmental exposure, Mutagenesis, 30 (2015) 763-770.
[7] B. An, T. Pan, J. Hu, Y. Pang, L. Huang, A.S.C. Chan, X. Li, J. Yan, The discovery of a potent and selective third-generation EGFR kinase inhibitor as a therapy for EGFR L858R/T790M double mutant non-small cell lung cancer, Eur J Med Chem, 183 (2019) 111709.
[8] J.P. Allison, Checkpoints, Cell, 162 (2015) 1202-1205.
[9] T. Schumacher, L. Bunse, S. Pusch, F. Sahm, B. Wiestler, J. Quandt, O. Menn, M. Osswald, I. Oezen, M. Ott, M. Keil, J. Balss, K. Rauschenbach, A.K. Grabowska, I. Vogler, J. Diekmann, N. Trautwein, S.B. Eichmuller, J. Okun, S. Stevanovic, A.B. Riemer, U. Sahin, M.A. Friese, P. Beckhove, A. von Deimling, W. Wick, M. Platten, A vaccine targeting mutant IDH1 induces antitumour immunity, Nature, 512 (2014) 324-327.
[10] B.M. Carreno, V. Magrini, M. Becker-Hapak, S. Kaabinejadian, J. Hundal, A.A. Petti, A. Ly, W.R. Lie, W.H. Hildebrand, E.R. Mardis, G.P. Linette, Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells, Science, 348 (2015) 803-808.
[11] T. Schlake, A. Thess, M. Fotin-Mleczek, K.J. Kallen, Developing mRNA-vaccine technologies, RNA Biol, 9 (2012) 1319-1330.
[12] S. Colloca, E. Barnes, A. Folgori, V. Ammendola, S. Capone, A. Cirillo, L. Siani, M. Naddeo, F. Grazioli, M.L. Esposito, M. Ambrosio, A. Sparacino, M. Bartiromo, A. Meola, K. Smith, A. Kurioka, G.A. O'Hara, K.J. Ewer, N. Anagnostou, C. Bliss, A.V. Hill, C. Traboni, P. Klenerman, R. Cortese, A. Nicosia, Vaccine vectors derived from a large collection of simian adenoviruses induce potent cellular immunity across multiple species, Sci Transl Med, 4 (2012) 115ra112.
[13] B. Zhang, J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M.C. Chambers, L.J. Zimmerman, K.F. Shaddox, S. Kim, S.R. Davies, S. Wang, P. Wang, C.R. Kinsinger, R.C. Rivers, H. Rodriguez, R.R. Townsend, M.J. Ellis, S.A. Carr, D.L. Tabb, R.J. Coffey, R.J. Slebos, D.C. Liebler, C. Nci, Proteogenomic characterization of human colon and rectal cancer, Nature, 513 (2014) 382-387.
[14] K.V. Ruggles, K. Krug, X. Wang, K.R. Clauser, J. Wang, S.H. Payne, D. Fenyo, B. Zhang, D.R. Mani, Methods, Tools and Current Perspectives in Proteogenomics, Mol Cell Proteomics, 16 (2017) 959-981.
[15] S. Apcher, C. Daskalogianni, R. Fahraeus, Pioneer translation products as an alternative source for MHC-I antigenic peptides, Mol Immunol, 68 (2015) 68-71.
[16] J. Crappe, E. Ndah, A. Koch, S. Steyaert, D. Gawron, S. De Keulenaer, E. De Meester, T. De Meyer, W. Van Criekinge, P. Van Damme, G. Menschaert, PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration, Nucleic Acids Res, 43 (2015) e29.
[17] G.S. Krasnov, A.A. Dmitriev, A.V. Kudryavtseva, A.V. Shargunov, D.S. Karpov, L.A. Uroshlev, N.V. Melnikova, V.M. Blinov, E.V. Poverennaya, A.I. Archakov, A.V. Lisitsa, E.A. Ponomarenko,

PPLine: An Automated Pipeline for SNP, SAP, and Splice Variant Detection in the Context of Proteogenomics, J Proteome Res, 14 (2015) 3729-3737.
[18] P. Jagtap, J. Goslinga, J.A. Kooren, T. McGowan, M.S. Wroblewski, S.L. Seymour, T.J. Griffin, A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies, Proteomics, 13 (2013) 1352-1357.
[19] B. Afsari, T. Guo, M. Considine, L. Florea, L.T. Kagohara, G.L. Stein-O'Brien, D. Kelley, E. Flam, K.D. Zambo, P.K. Ha, D. Geman, M.F. Ochs, J.A. Califano, D.A. Gaykalova, A.V. Favorov, E.J. Fertig, Splice Expression Variation Analysis (SEVA) for inter-tumor heterogeneity of gene isoform usage in cancer, Bioinformatics, 34 (2018) 1859-1867.
[20] B. Wen, S. Xu, G.M. Sheynkman, Q. Feng, L. Lin, Q. Wang, X. Xu, J. Wang, S. Liu, sapFinder: an R/Bioconductor package for detection of variant peptides in shotgun proteomics experiments, Bioinformatics, 30 (2014) 3136-3138.
[21] F. Zickmann, B.Y. Renard, MSProGene: integrative proteogenomics beyond six-frames and single nucleotide polymorphisms, Bioinformatics, 31 (2015) i106-115.
[22] J. Hedegaard, K. Thorsen, M.K. Lund, A.M. Hein, S.J. Hamilton-Dutoit, S. Vang, I. Nordentoft, K. Birkenkamp-Demtroder, M. Kruhoffer, H. Hager, B. Knudsen, C.L. Andersen, K.D. Sorensen, J.S. Pedersen, T.F. Orntoft, L. Dyrskjot, Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue, PLoS One, 9 (2014) e98187.
[23] V. Tiedje, S. Ting, T. Herold, S. Synoracki, S. Latteyer, L.C. Moeller, D. Zwanziger, M. Stuschke, D. Fuehrer, K.W. Schmid, NGS based identification of mutational hotspots for targeted therapy in anaplastic thyroid carcinoma, Oncotarget, 8 (2017) 42613-42620.
[24] V. Gorshkov, T. Verano-Braga, F. Kjeldsen, SuperQuant: A Data Processing Approach to Increase Quantitative Proteome Coverage, Anal Chem, 87 (2015) 6319-6327.
[25] I.F. do Valle, E. Giampieri, G. Simonetti, A. Padella, M. Manfrini, A. Ferrari, C. Papayannidis, I. Zironi, M. Garonzi, S. Bernardi, M. Delledonne, G. Martinelli, D. Remondini, G. Castellani, Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data, BMC Bioinformatics, 17 (2016) 341.
[26] J.P. Blaydes, M.G. Luciani, S. Pospisilova, H.M. Ball, B. Vojtesek, T.R. Hupp, Stoichiometric phosphorylation of human p53 at Ser315 stimulates p53-dependent transcription, J Biol Chem, 276 (2001) 4699-4708.
[27] J.P. Blaydes, T.R. Hupp, DNA damage triggers DRB-resistant phosphorylation of human p53 at the CK2 site, Oncogene, 17 (1998) 1045-1052.
[28] S.P. Shah, A. Roth, R. Goya, A. Oloumi, G. Ha, Y. Zhao, G. Turashvili, J. Ding, K. Tse, G. Haffari, A. Bashashati, L.M. Prentice, J. Khattra, A. Burleigh, D. Yap, V. Bernard, A. McPherson, K. Shumansky, A. Crisan, R. Giuliany, A. Heravi-Moussavi, J. Rosner, D. Lai, I. Birol, R. Varhol, A. Tam, N. Dhalla, T. Zeng, K. Ma, S.K. Chan, M. Griffith, A. Moradian, S.W. Cheng, G.B. Morin, P. Watson, K. Gelmon, S. Chia, S.F. Chin, C. Curtis, O.M. Rueda, P.D. Pharoah, S. Damaraju, J. Mackey, K. Hoon, T. Harkins, V. Tadigotla, M. Sigaroudinia, P. Gascard, T. Tlsty, J.F. Costello, I.M. Meyer, C.J. Eaves, W.W. Wasserman, S. Jones, D. Huntsman, M. Hirst, C. Caldas, M.A. Marra, S. Aparicio, The clonal and mutational evolution spectrum of primary triple-negative breast cancers, Nature, 486 (2012) 395-399.
[29] R.D. Morin, M. Mendez-Lago, A.J. Mungall, R. Goya, K.L. Mungall, R.D. Corbett, N.A. Johnson, T.M. Severson, R. Chiu, M. Field, S. Jackman, M. Krzywinski, D.W. Scott, D.L. Trinh, J. Tamura-Wells, S. Li, M.R. Firme, S. Rogic, M. Griffith, S. Chan, O. Yakovenko, I.M. Meyer, E.Y. Zhao, D. Smailus, M. Moksa, S. Chittaranjan, L. Rimsza, A. Brooks-Wilson, J.J. Spinelli, S. Ben-Neriah, B. Meissner, B. Woolcock, M. Boyle, H. McDonald, A. Tam, Y. Zhao, A. Delaney, T. Zeng, K. Tse, Y. Butterfield, I. Birol, R. Holt, J. Schein, D.E. Horsman, R. Moore, S.J. Jones, J.M. Connors, M. Hirst, R.D. Gascoyne, M.A. Marra, Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma, Nature, 476 (2011) 298-303.

[30] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, Nat Methods, 6 (2009) 359-362.
[31] B.C. Collins, C.L. Hunter, Y. Liu, B. Schilling, G. Rosenberger, S.L. Bader, D.W. Chan, B.W. Gibson, A.C. Gingras, J.M. Held, M. Hirayama-Kurogi, G. Hou, C. Krisp, B. Larsen, L. Lin, S. Liu, M.P. Molloy, R.L. Moritz, S. Ohtsuki, R. Schlapbach, N. Selevsek, S.N. Thomas, S.C. Tzeng, H. Zhang, R. Aebersold, Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry, Nat Commun, 8 (2017) 291.
[32] L. Ding, M. Kim, K.L. Kanchi, N.D. Dees, C. Lu, M. Griffith, D. Fenstermacher, H. Sung, C.A. Miller, B. Goetz, M.C. Wendl, O. Griffith, L.A. Cornelius, G.P. Linette, J.F. McMichael, V.K. Sondak, R.C. Fields, T.J. Ley, J.J. Mule, R.K. Wilson, J.S. Weber, Clonal architectures and driver mutations in metastatic melanomas, PLoS One, 9 (2014) e111153.
[33] P.A. Ott, Z. Hu, D.B. Keskin, S.A. Shukla, J. Sun, D.J. Bozym, W. Zhang, A. Luoma, A. Giobbie-Hurder, L. Peter, C. Chen, O. Olive, T.A. Carter, S. Li, D.J. Lieb, T. Eisenhaure, E. Gjini, J. Stevens, W.J. Lane, I. Javeri, K. Nellaiappan, A.M. Salazar, H. Daley, M. Seaman, E.I. Buchbinder, C.H. Yoon, M. Harden, N. Lennon, S. Gabriel, S.J. Rodig, D.H. Barouch, J.C. Aster, G. Getz, K. Wucherpfennig, D. Neuberg, J. Ritz, E.S. Lander, E.F. Fritsch, N. Hacohen, C.J. Wu, An immunogenic personal neoantigen vaccine for patients with melanoma, Nature, 547 (2017) 217-221.
[34] D.P. Lane, Cancer. p53, guardian of the genome, Nature, 358 (1992) 15-16.
[35] K. Edlund, O. Larsson, A. Ameur, I. Bunikis, U. Gyllensten, B. Leroy, M. Sundstrom, P. Micke, J. Botling, T. Soussi, Data-driven unbiased curation of the TP53 tumor suppressor gene mutation database and validation by ultradeep sequencing of human tumors, Proc Natl Acad Sci U S A, 109 (2012) 9551-9556.
[36] M. Overholtzer, P.H. Rao, R. Favis, X.Y. Lu, M.B. Elowitz, F. Barany, M. Ladanyi, R. Gorlick, A.J. Levine, The presence of p53 mutations in human osteosarcomas correlates with high levels of genomic instability, Proc Natl Acad Sci U S A, 100 (2003) 11547-11552.
[37] M.S. Lawrence, P. Stojanov, P. Polak, G.V. Kryukov, K. Cibulskis, A. Sivachenko, S.L. Carter, C. Stewart, C.H. Mermel, S.A. Roberts, A. Kiezun, P.S. Hammerman, A. McKenna, Y. Drier, L. Zou, A.H. Ramos, T.J. Pugh, N. Stransky, E. Helman, J. Kim, C. Sougnez, L. Ambrogio, E. Nickerson, E. Shefler, M.L. Cortes, D. Auclair, G. Saksena, D. Voet, M. Noble, D. DiCara, P. Lin, L. Lichtenstein, D.I. Heiman, T. Fennell, M. Imielinski, B. Hernandez, E. Hodis, S. Baca, A.M. Dulak, J. Lohr, D.A. Landau, C.J. Wu, J. Melendez-Zajgla, A. Hidalgo-Miranda, A. Koren, S.A. McCarroll, J. Mora, B. Crompton, R. Onofrio, M. Parkin, W. Winckler, K. Ardlie, S.B. Gabriel, C.W.M. Roberts, J.A. Biegel, K. Stegmaier, A.J. Bass, L.A. Garraway, M. Meyerson, T.R. Golub, D.A. Gordenin, S. Sunyaev, E.S. Lander, G. Getz, Mutational heterogeneity in cancer and the search for new cancer-associated genes, Nature, 499 (2013) 214-218.
[38] J.M.J. Weaver, C.S. Ross-Innes, N. Shannon, A.G. Lynch, T. Forshew, M. Barbera, M. Murtaza, C.J. Ong, P. Lao-Sirieix, M.J. Dunning, L. Smith, M.L. Smith, C.L. Anderson, B. Carvalho, M. O'Donovan, T.J. Underwood, A.P. May, N. Grehan, R. Hardwick, J. Davies, A. Oloumi, S. Aparicio, C. Caldas, M.D. Eldridge, P.A.W. Edwards, N. Rosenfeld, S. Tavare, R.C. Fitzgerald, O. consortium, Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis, Nat Genet, 46 (2014) 837-843.
[39] E. Garcia, A. Hayden, C. Birts, E. Britton, A. Cowie, K. Pickard, M. Mellone, C. Choh, M. Derouet, P. Duriez, F. Noble, M.J. White, J.N. Primrose, J.C. Strefford, M. Rose-Zerilli, G.J. Thomas, Y. Ang, A.D. Sharrocks, R.C. Fitzgerald, T.J. Underwood, O. consortium, Authentication and characterisation of a new oesophageal adenocarcinoma cell line: MFD-1, Sci Rep, 6 (2016) 32417.
[40] M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system, Bioinformatics, 32 (2016) 511-517.
[41] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system, Nat Protoc, 8 (2013) 2281-2308.

[42] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S.L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol, 14 (2013) R36.
[43] D.C. Koboldt, Q. Zhang, D.E. Larson, D. Shen, M.D. McLellan, L. Lin, C.A. Miller, E.R. Mardis, L. Ding, R.K. Wilson, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing, Genome Res, 22 (2012) 568-576.
[44] T.S. Batth, C. Francavilla, J.V. Olsen, Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics, J Proteome Res, 13 (2014) 6176-6186.
[45] L.C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, Mol Cell Proteomics, 11 (2012) O111 016717.
[46] L. Way, J. Faktor, P. Dvorakova, J. Nicholson, B. Vojtesek, D. Graham, K.L. Ball, T. Hupp, Rearrangement of mitochondrial pyruvate dehydrogenase subunit dihydrolipoamide dehydrogenase protein-protein interactions by the MDM2 ligand nutlin-3, Proteomics, 16 (2016) 2327-2344.
[47] F. Zavadil Kokas, J. Faktor, B. Vojtesek, Cooperation of Genomic, Transcriptomics and Proteomic Methods in the Detection of Mutated Proteins, Klin Onkol, 32 (2019) 78-84.
[48] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res, 47 (2019) D607-D613.

Supplementary table 1A

Click here to download Supplementary Material (for online publication): Supp Table 1A.xlsx

Chromosome	Region	Type	Reference	Allele	Reference allele	Length	Zygosity	Count	Coverage
1	783029	SNV	C	A	No	1	Heterozygo	2	22
1	783029	SNV	C	C	Yes	1	Heterozygo	20	22
1	824115	SNV	A	C	No	1	Heterozygo	7	10
1	824142	SNV	C	T	No	1	Heterozygo	8	11
1	824145..82	MNV	CT	TG	No	2	Heterozygo	8	11
1	877992	SNV	C	A	No	1	Heterozygo	2	12
1	877992	SNV	C	C	Yes	1	Heterozygo	10	12
1	880101	SNV	C	A	No	1	Heterozygo	2	12
1	880101	SNV	C	C	Yes	1	Heterozygo	10	12
1	881679	SNV	G	T	No	1	Heterozygo	4	46
1	881679	SNV	G	G	Yes	1	Heterozygo	42	46
1	884091	SNV	C	G	No	1	Heterozygo	23	63
1	884091	SNV	C	C	Yes	1	Heterozygo	39	63
1	884101	SNV	A	C	No	1	Heterozygo	30	56
1	884101	SNV	A	A	Yes	1	Heterozygo	26	56
1	894642	Deletion	C	-	No	1	Heterozygo	2	29
1	898085	SNV	T	C	No	1	Heterozygo	2	40
1	898085	SNV	T	T	Yes	1	Heterozygo	38	40
1	906643	SNV	G	A	No	1	Heterozygo	2	37
1	906643	SNV	G	G	Yes	1	Heterozygo	34	37
1	908654	SNV	C	A	No	1	Heterozygo	3	51
1	908654	SNV	C	C	Yes	1	Heterozygo	48	51
1	915885	SNV	G	A	No	1	Heterozygo	58	84
1	915885	SNV	G	G	Yes	1	Heterozygo	26	84
1	977191..97	MNV	CGG	TCT	No	3	Heterozygo	3	13
1	977191..97	MNV	CGG	CGG	Yes	3	Heterozygo	10	13
1	977202	SNV	T	C	No	1	Heterozygo	2	13
1	977202	SNV	T	T	Yes	1	Heterozygo	11	13
1	978581	SNV	G	C	No	1	Heterozygo	2	40
1	978581	SNV	G	G	Yes	1	Heterozygo	38	40
1	983673	SNV	C	A	No	1	Heterozygo	4	65
1	983673	SNV	C	C	Yes	1	Heterozygo	61	65
1	984267	SNV	G	A	No	1	Heterozygo	2	13
1	984267	SNV	G	G	Yes	1	Heterozygo	11	13
1	990380	SNV	C	T	No	1	Heterozygo	16	24
1	990380	SNV	C	C	Yes	1	Heterozygo	8	24
1	1133117	SNV	C	A	No	1	Heterozygo	2	14
1	1133117	SNV	C	C	Yes	1	Heterozygo	12	14
1	1133120	SNV	G	T	No	1	Heterozygo	2	14
1	1133120	SNV	G	G	Yes	1	Heterozygo	12	14

Supplementary table 1B

Click here to download Supplementary Material (for online publication): Supp Table 1B.xlsx

Chromosome	Region	Type	Reference	Allele	Reference allele	Length
1	12887549	SNV	T	C	No	1
1	13036587	SNV	C	T	No	1
1	16383742	SNV	C	G	No	1
1	28268833^282	Insertion	-	G	No	1
1	79128570^791	Insertion	-	T	No	1
1	152187562	SNV	A	C	No	1
1	161514542	SNV	A	C	No	1
1	248458717	SNV	C	G	No	1
1	248651927	SNV	A	G	No	1
1	248722722	SNV	T	C	No	1
1	248801610..24	MNV	GC	AT	No	2
2	24402127^244	Insertion	-	AA	No	2
2	73675227^736	Insertion	-	CTC	No	3
2	95847041..958	Deletion	GCG	-	No	3
2	120194651^12	Insertion	-	GTGTGC	No	6
2	120704166	SNV	G	T	No	1
2	130738163	SNV	G	A	No	1
2	169830328	SNV	A	G	No	1
2	172182386	SNV	C	G	No	1
2	178494173^17	Insertion	-	GGA	No	3
3	33879770	SNV	G	A	No	1
3	64619524	SNV	G	A	No	1
3	96336047	SNV	G	A	No	1
3	96336057	SNV	T	G	No	1
3	96336077	SNV	C	T	No	1
3	96336080	SNV	G	A	No	1
3	96336084..963	MNV	AA	GG	No	2
3	96336087	Deletion	G	-	No	1
3	96336090^963	Insertion	-	G	No	1
3	96336098	SNV	C	T	No	1
3	96336100..963	MNV	GAA	ATG	No	3
3	96336111	SNV	G	T	No	1
3	96336113	SNV	G	A	No	1
3	96336116	SNV	A	T	No	1
3	97983391	SNV	C	T	No	1

Supplementary table 2

Click here to download Supplementary Material (for online publication): Supp Table 2.xlsx

Chromosome	Region	Type	Reference	Allele	Reference allele	Length
1	12887549	SNV	T	C	No	1
1	13036587	SNV	C	T	No	1
1	16383742	SNV	C	G	No	1
1	28268833^282	Insertion	-	G	No	1
1	79128570^791	Insertion	-	T	No	1
1	152187562	SNV	A	C	No	1
1	161514542	SNV	A	C	No	1
1	248458717	SNV	C	G	No	1
1	248651927	SNV	A	G	No	1
1	248722722	SNV	T	C	No	1
1	248801610..24	MNV	GC	AT	No	2
2	24402127^244	Insertion	-	AA	No	2
2	73675227^736	Insertion	-	CTC	No	3
2	95847041..958	Deletion	GCG	-	No	3
2	120194651^12	Insertion	-	GTGTGC	No	6
2	120704166	SNV	G	T	No	1
2	130738163	SNV	G	A	No	1
2	169830328	SNV	A	G	No	1
2	172182386	SNV	C	G	No	1
2	178494173^17	Insertion	-	GGA	No	3
3	33879770	SNV	G	A	No	1
3	64619524	SNV	G	A	No	1
3	96336047	SNV	G	A	No	1
3	96336057	SNV	T	G	No	1
3	96336077	SNV	C	T	No	1
3	96336080	SNV	G	A	No	1
3	96336084..963	MNV	AA	GG	No	2
3	96336087	Deletion	G	-	No	1
3	96336090^963	Insertion	-	G	No	1
3	96336098	SNV	C	T	No	1
3	96336100..963	MNV	GAA	ATG	No	3
3	96336111	SNV	G	T	No	1
3	96336113	SNV	G	A	No	1
3	96336116	SNV	A	T	No	1
3	97983391	SNV	C	T	No	1
3	123457893	SNV	G	A	No	1
3	184429133^18	Insertion	-	TCC	No	3
3	195512107	SNV	T	A	No	1

Supplementary table 3

Click here to download Supplementary Material (for online publication): Supp Table 3.xlsx

Chromosor	Region	Type	Reference	Allele	Reference allele	Length	Zygosity	Count
3	52027853^	Insertion	-	CCTTGG	No	6	Homozygo	4247
3	40503520^	Insertion	-	CTGCTG	No	6	Heterozygo	667
6	1.6E+08	SNV	T	C	No	1	Heterozygo	591
12	7080212	SNV	T	C	No	1	Homozygo	559
X	1.36E+08	SNV	G	T	No	1	Heterozygo	535
2	232325415	Deletion	TCA	-	No	3	Heterozygo	475
1	6257784^6	Insertion	-	T	No	1	Heterozygo	442
8	86126827^	Insertion	-	AACATT	No	6	Homozygo	424
8	1.45E+08	SNV	C	T	No	1	Heterozygo	407
6	26157119^	Insertion	-	A	No	1	Heterozygo	372
11	86519181	SNV	G	T	No	1	Heterozygo	368
12	53690045	SNV	C	T	No	1	Heterozygo	344
19	51850290	SNV	G	A	No	1	Homozygo	334
20	21314624	SNV	A	G	No	1	Heterozygo	321
20	25271155	SNV	G	C	No	1	Heterozygo	310
12	7077684	SNV	G	A	No	1	Heterozygo	305
8	125528031	Deletion	TTTAAAAA	-	No	98	Heterozygo	277
1	156565049	Insertion	-	AC	No	2	Homozygo	275
X	12994411	SNV	G	A	No	1	Heterozygo	270
6	29912856	SNV	A	T	No	1	Heterozygo	266
6	29911092	SNV	G	T	No	1	Heterozygo	258
7	80433462	SNV	C	T	No	1	Homozygo	255
6	29911114.	MNV	GG	AC	No	2	Heterozygo	246
6	29911119	SNV	G	T	No	1	Heterozygo	243
2	26477125^	Insertion	-	ACT	No	3	Homozygo	237
6	29911154	SNV	C	A	No	1	Heterozygo	236
6	29911207	SNV	G	A	No	1	Heterozygo	233
6	29911901	SNV	C	G	No	1	Heterozygo	232
6	29911056	SNV	A	G	No	1	Heterozygo	230
6	29911228	SNV	A	T	No	1	Heterozygo	225
14	21679995^	Insertion	-	T	No	1	Heterozygo	223
9	1.31E+08	SNV	G	A	No	1	Heterozygo	222
1	2.37E+08	SNV	A	G	No	1	Heterozygo	219
8	56982391.	Replaceme	AA	T	No	2	Homozygo	213
6	29911198	SNV	T	C	No	1	Heterozygo	213
14	102549379	Insertion	-	T	No	1	Heterozygo	210
4	1.91E+08	Deletion	T	-	No	1	Heterozygo	209
17	17039562.	Deletion	CAG	-	No	3	Homozygo	206
6	29911928	SNV	C	G	No	1	Heterozygo	201
6	29911930.	MNV	CA	TG	No	2	Heterozygo	201
1	2.29E+08	SNV	C	G	No	1	Heterozygo	201
1	21091915	SNV	C	T	No	1	Heterozygo	195
2	55898372.	Deletion	AATCCCAA	-	No	86	Homozygo	185
4	1.09E+08	SNV	G	A	No	1	Homozygo	180
5	1.38E+08	SNV	T	G	No	1	Heterozygo	178
13	31037379^	Insertion	-	T	No	1	Heterozygo	176

Supplementary table 4

Click here to download Supplementary Material (for online publication): Supp Table 4.xlsx

Chromos	Region	Type	Zygosity	Gene Cards	ENSEMBL	gene_name (Ho
1	888659	SNV	Homozygous	NOC2L	ENSG0000018	NOC2L
1	979748	SNV	Heterozygous	AGRN	ENSG0000018	AGRN
1	1E+06	SNV	Homozygous	CPSF3L	ENSG0000012	CPSF3L
1	1E+06	SNV	Heterozygous	ATAD3B	ENSG0000016	ATAD3B
1	1E+06	SNV	Heterozygous	ATAD3A	ENSG0000019	ATAD3A
1	1647893	Insertion	Unknown	CDK11A , RP1-28	ENSG0000000	CDK11A, RP1-28
1	2E+06	SNV	Unknown	CDK11A , RP1-28	ENSG0000000	CDK11A
1	2E+06	SNV	Unknown	CDK11A , RP1-28	ENSG0000000	CDK11A, RP1-28
1	2E+06	SNV	Unknown	CDK11A , RP1-28	ENSG0000000	CDK11A, RP1-28
1	2E+06	SNV	Homozygous	NADK	ENSG0000000	NADK
1	4E+06	SNV	Homozygous	C1orf174	ENSG0000019	C1orf174
1	7E+06	SNV	Homozygous	THAP3	ENSG0000004	THAP3
1	1E+07	SNV	Unknown	MTOR	ENSG0000019	MTOR
1	1E+07	SNV	Unknown	MIIP	ENSG0000011	MIIP
1	2E+07	SNV	Homozygous	RP1-43E13.2 , EM	ENSG0000023	EMC1
1	2E+07	SNV	Homozygous	DDOST	ENSG0000024	DDOST
1	2E+07	SNV	Heterozygous	HP1BP3	ENSG0000012	HP1BP3
1	2E+07	SNV	Unknown	EIF4G3	ENSG0000007	EIF4G3
1	2E+07	SNV	Homozygous	ID3	ENSG0000011	ID3
1	2E+07	SNV	Unknown	NIPAL3	ENSG0000000	NIPAL3
1	3E+07	SNV	Homozygous	RPS6KA1	ENSG0000011	RPS6KA1
1	3E+07	SNV	Heterozygous	RPS6KA1	ENSG0000011	RPS6KA1
1	3E+07	SNV	Homozygous	GPN2	ENSG0000014	GPN2
1	3E+07	SNV	Homozygous	TRNP1	ENSG0000025	TRNP1
1	3E+07	SNV	Heterozygous	PPP1R8	ENSG0000011	PPP1R8
1	3E+07	SNV	Homozygous	SRSF4	ENSG0000011	SRSF4
1	3E+07	SNV	Homozygous	SRSF4	ENSG0000011	SRSF4
1	3E+07	SNV	Homozygous	SRSF4	ENSG0000011	SRSF4
1	3E+07	SNV	Homozygous	MECR	ENSG0000011	MECR
1	3190588	Insertion	Heterozygous	SERINC2	ENSG0000016	SERINC2
1	3E+07	SNV	Heterozygous	ZBTB8OS	ENSG0000017	ZBTB8OS
1	4E+07	SNV	Homozygous	ZMYM1	ENSG0000019	ZMYM1
1	4E+07	SNV	Homozygous	CLSPN	ENSG0000009	CLSPN
1	4E+07	SNV	Unknown	CLSPN	ENSG0000009	CLSPN
1	4E+07	SNV	Heterozygous	ADPRHL2	ENSG0000011	ADPRHL2
1	4E+07	SNV	Heterozygous	THRAP3	ENSG0000005	THRAP3
1	4E+07	SNV	Unknown	MACF1	ENSG0000012	MACF1
1	4E+07	SNV	Homozygous	CAP1	ENSG0000013	CAP1
1	4E+07	SNV	Homozygous	CAP1	ENSG0000013	CAP1
1	4E+07	SNV	Homozygous	CAP1	ENSG0000013	CAP1
1	4E+07	SNV	Homozygous	CAP1	ENSG0000013	CAP1
1	4E+07	SNV	Homozygous	CAP1	ENSG0000013	CAP1

Supplementary table 5

Click here to download Supplementary Material (for online publication): Supp Table 5.xlsx

Chromosome	Region	Type	Zygosity	Gene Cards	ENSEMBL	gene_name (
1	1647893^164	Insertion	Unknown	CDK11A , RP1	ENSG000000(	CDK11A, RP1-
1	1650787	SNV	Unknown	CDK11A , RP1	ENSG000000(	CDK11A
1	1650797	SNV	Unknown	CDK11A , RP1	ENSG000000(	CDK11A, RP1-
1	1650832	SNV	Unknown	CDK11A , RP1	ENSG000000(	CDK11A, RP1-
1	11186751	SNV	Unknown	MTOR	ENSG000001!	MTOR
1	12082926	SNV	Unknown	MIIP	ENSG000001:	MIIP
1	21267993	SNV	Unknown	EIF4G3	ENSG000000	EIF4G3
1	24785393	SNV	Unknown	NIPAL3	ENSG000000(	NIPAL3
1	36226120	SNV	Unknown	CLSPN	ENSG000000!	CLSPN
1	39914363	SNV	Unknown	MACF1	ENSG000001:	MACF1
1	42657270	SNV	Unknown	FOXJ3	ENSG000001!	FOXJ3
1	47799639	SNV	Unknown	CMPK1	ENSG000001(	CMPK1
1	52861871	SNV	Unknown	ORC1	ENSG000000!	ORC1
1	98144726	SNV	Unknown	DPYD	ENSG000001!	DPYD
1	107599918	SNV	Unknown	PRMT6	ENSG000001!	PRMT6
1	112308953	SNV	Unknown	DDX20	ENSG000000(	DDX20
1	112309123	SNV	Unknown	DDX20	ENSG000000(	DDX20
1	112309331	SNV	Unknown	DDX20	ENSG000000(	DDX20
1	117529458	SNV	Unknown	PTGFRN	ENSG000001:	PTGFRN
1	151315287	SNV	Unknown	RFX5	ENSG000001-	RFX5
1	153615820	SNV	Unknown	CHTOP	ENSG000001(	CHTOP
1	156640156	SNV	Unknown	NES	ENSG000001:	NES
1	159002377	SNV	Unknown	IFI16	ENSG000001(	IFI16
1	159002389	SNV	Unknown	IFI16	ENSG000001(	IFI16
1	165619079	SNV	Unknown	MGST3	ENSG000001-	MGST3
1	171486912	SNV	Unknown	PRRC2C	ENSG000001:	PRRC2C
1	179852074	SNV	Unknown	TOR1AIP1	ENSG000001-	TOR1AIP1
1	179989742	SNV	Unknown	CEP350	ENSG000001:	CEP350
1	186363119	SNV	Unknown	C1orf27	ENSG000001!	C1orf27
1	210004199	SNV	Unknown	DIEXF	ENSG000001:	DIEXF
1	210415570	SNV	Unknown	SERTAD4	ENSG000000!	SERTAD4
1	226054333	SNV	Unknown	TMEM63A	ENSG000001!	TMEM63A
1	227216775	SNV	Unknown	CDC42BPA	ENSG000001-	CDC42BPA
1	227935444	SNV	Unknown	SNAP47	ENSG000001-	SNAP47
1	230415148	SNV	Unknown	GALNT2	ENSG000001-	GALNT2
1	236175327	SNV	Unknown	NID1	ENSG000001:	NID1
1	236700807	SNV	Unknown	LGALS8	ENSG000001:	LGALS8
1	236700857	SNV	Unknown	LGALS8	ENSG000001:	LGALS8
1	236706862	SNV	Unknown	LGALS8 , RP1	ENSG000001:	LGALS8
1	237048500	SNV	Unknown	MTR	ENSG000001:	MTR

Supplementary table 6

Click here to download Supplementary Material (for online publication): Supp Table 6.xlsx

ID: CDK11A , RP1-283E3.8
ID: MIIP
ID: EIF4G3
ID: NIPAL3
ID: CLSPN
ID: MACF1
ID: FOXJ3
ID: CMPK1
ID: ORC1
ID: DPYD
ID: PRMT6
ID: DDX20
ID: PTGFRN
ID: RFX5
ID: CHTOP
ID: NES
ID: IFI16
ID: MGST3
ID: PRRC2C
ID: TOR1AIP1
ID: CEP350
ID: C1orf27
ID: DIEXF
ID: SERTAD4
ID: TMEM63A
ID: SNAP47
ID: GALNT2
ID: NID1
ID: LGALS8
ID: LGALS8 , RP11-385F5.4
ID: MTR
ID: SDCCAG8
ID: ACP1
ID: TRAPPC12
ID: NBAS
ID: SLC5A6
ID: EML4
ID: PIGF
ID: RAD18
ID: MAP4
ID: SHQ1

Supplementary table 7

Click here to download Supplementary Material (for online publication): Supp Table 7.xlsx

Row	Index	Peak Name	m/z	Ret. Time
598	598	sp|O1477	N/A	N/A
360	360	tr|B4DGU	N/A	N/A
875	875	sp|O4377	N/A	N/A
785	785	tr|C9JDE9	N/A	N/A
729	729	sp|P04259	N/A	N/A
634	634	sp|P41252	N/A	N/A
937	937	sp|O4327	N/A	N/A
513	513	sp|P35908	N/A	N/A
852	852	tr|D6REX3	N/A	N/A
291	291	sp|Q9HAV	N/A	N/A
762	762	tr|X6R700	N/A	N/A
558	558	sp|Q9282	N/A	N/A
680	680	sp|P00387	N/A	N/A
887	887	sp|Q9995	N/A	N/A
601	601	tr|A0A087	N/A	N/A
739	739	sp|P23229	N/A	N/A
735	735	sp|P62873	N/A	N/A
813	813	sp|P62304	N/A	N/A
888	888	sp|Q9289	N/A	N/A
728	728	sp|Q9NZB	N/A	N/A
796	796	tr|I3L1P8	N/A	N/A
395	395	sp|P11216	N/A	N/A
826	826	tr|F5H895	N/A	N/A
611	611	tr|J3KN36	N/A	N/A
769	769	sp|Q9BRR	N/A	N/A
948	948	sp|P00533	N/A	N/A
77	77	sp|P13804	N/A	N/A
910	910	tr|A0A087	N/A	N/A
772	772	sp|Q1573	N/A	N/A
569	569	sp|Q9NYU	N/A	N/A
430	430	sp|P11166	N/A	N/A
312	312	sp|P53007	N/A	N/A
671	671	tr|F5GZ78	N/A	N/A
509	509	tr|Q5VV8	N/A	N/A
916	916	tr|C9JQD4	N/A	N/A
859	859	sp|Q9P0S	N/A	N/A
899	899	sp|Q9NX5	N/A	N/A
507	507	sp|P07711	N/A	N/A
263	263	sp|P04792	N/A	N/A
448	448	sp|Q9NQC	N/A	N/A
466	466	tr|A0A087	N/A	N/A
894	894	sp|Q96BR	N/A	N/A
871	871	sp|P11047	N/A	N/A
200	200	sp|Q96FQ	N/A	N/A
624	624	sp|Q6DRA	N/A	N/A
615	615	tr|Q5QPL9	N/A	N/A
811	811	sp|Q0180	N/A	N/A
831	831	sp|P49006	N/A	N/A
106	106	sp|P26038	N/A	N/A

Supplementary table 8A

Click here to download Supplementary Material (for online publication): Supp Table 8A.xlsx

N	Unused	Total	%Cov	%Cov(50)	%Cov(95)	Accession	Name	Species	Peptides(9
1	406.2	406.2	67.55	56.02	52.54	sp|Q1514	Plectin OS=	HUMAN	221
2	257.09	257.09	70.87	62.38	56.54	sp|Q0966	Neuroblast	HUMAN	181
3	256.77	256.77	78.17	74.37	73.1	sp|O7536	Filamin-B (	HUMAN	171
4	230.75	247.34	73.33	68.87	67.09	sp|P21333	Filamin-A (	HUMAN	182
4	0	245.86	72.54	68.85	67.05	tr|Q5HY54	Filamin-A (	HUMAN	181
4	0	245.66	73.09	68.93	67.14	tr|Q60FE5	Filamin A (	HUMAN	181
5	215.25	215.25	59.5	51.62	46.71	sp|P78527	DNA-depe	HUMAN	148
6	204.27	204.27	66.79	58.37	55.87	sp|P35579	Myosin-9 (	HUMAN	150
7	198.26	198.26	68.37	62.38	59.3	sp|Q1381	Spectrin al	HUMAN	117
8	192.38	192.38	76.36	73.55	69.37	sp|Q0061	Clathrin he	HUMAN	136
8	0	192.31	76.18	73.38	69.21	tr|A0A087	Clathrin he	HUMAN	136
9	182.7	182.7	65.87	62.29	58.26	sp|P49327	Fatty acid	HUMAN	115
10	181.15	181.15	65.57	60.53	56.68	sp|Q0108	Spectrin be	HUMAN	109
11	169.99	169.99	44.19	38.53	34.55	sp|Q1420	Cytoplasm	HUMAN	112
12	168.34	168.34	69.22	63.2	61.24	sp|Q9Y49	Talin-1 OS=	HUMAN	108
13	160.8	160.8	86.94	85.59	85.59	sp|P07437	Tubulin be	HUMAN	163
14	160.5	160.5	88.36	87.46	87.16	sp|P04406	Glyceralde	HUMAN	212
15	158.33	158.33	73.62	68.78	68.65	sp|P08238	Heat shock	HUMAN	172
16	155.48	155.48	84.75	80.79	78.53	sp|P14618	Pyruvate k	HUMAN	182
17	147.96	147.96	88.8	85.33	79.73	sp|P63261	Actin, cyto	HUMAN	230
18	141.38	141.38	62.92	56.55	52.25	sp|O7564	U5 small n	HUMAN	83
19	138.96	139.38	83.42	80.46	79.14	sp|O4370	Alpha-actir	HUMAN	117
20	137.02	137.02	76.81	72.49	72.49	sp|P13639	Elongation	HUMAN	125
21	136.04	136.04	80.34	78.64	73.53	sp|P11142	Heat shock	HUMAN	147
22	135.87	135.87	85.84	82.19	79.83	sp|P0867(	Vimentin C	HUMAN	164
23	135.22	135.22	57.28	47.51	43.02	sp|Q9261	Translatior	HUMAN	81
24	131.74	131.74	67.13	62.76	59.06	sp|P07814	Bifunctiona	HUMAN	78
25	128.74	128.74	90.02	86.92	86.7	sp|P68363	Tubulin alp	HUMAN	134
26	126.96	127.17	83.41	80.88	80.88	sp|P06733	Alpha-enol	HUMAN	156
27	125.38	125.38	64.78	60.9	57.25	sp|P42704	Leucine-ric	HUMAN	75
28	121.1	121.1	78.91	76.8	74.69	sp|P55072	Transitiona	HUMAN	104
29	120.09	120.09	57.35	54.02	49.35	tr|F8VPD4	CAD protei	HUMAN	72
29	0	119.4	56	52.49	47.96	sp|P27708	CAD protei	HUMAN	72
30	116.17	116.17	52.85	45.87	42.18	sp|Q6P2Q	Pre-mRNA	HUMAN	71
31	115.23	115.23	42.02	34.73	30.64	sp|Q7Z6Z7	E3 ubiquiti	HUMAN	83
32	112.71	112.71	74.89	73.38	70.35	sp|P68104	Elongation	HUMAN	132
33	112.6	112.6	74.86	73.25	71.27	sp|P22314	Ubiquitin-l	HUMAN	80
34	106.52	106.52	66.22	61.57	59.21	sp|Q0821	ATP-depen	HUMAN	77
35	105.35	132.55	53.8	46.71	44.03	sp|P35580	Myosin-10	HUMAN	88
36	105.29	105.29	75.39	69.72	65.91	sp|P55060	Exportin-2	HUMAN	86
37	104.15	104.64	80.02	76.48	72.72	sp|Q1497	Importin s	HUMAN	79
38	102.98	102.98	54.68	48.7	46.59	sp|P46940	Ras GTPas	HUMAN	55
39	99.72	99.72	35.4	25.95	22.05	sp|Q5T4S7	E3 ubiquiti	HUMAN	72
40	99.68	99.68	75	66.53	61.44	sp|Q1469	Neutral alp	HUMAN	57
41	99.58	99.58	64.73	47.88	47.88	sp|Q0083	Heterogen	HUMAN	79
42	99.47	99.47	71.65	66.64	63.26	sp|O0041(	Importin-5	HUMAN	62

Supplementary table 8B

Click here to download Supplementary Material (for online publication): Supp Table 8B.xlsx

N	Unused	Total	%Cov	%Cov(50)	%Cov(95)	Accession	Name	Species	Peptides(9
1	391.83	391.83	65.48	55.36	49.25	sp|Q1514	Plectin OS=	HUMAN	214
2	258.4	258.4	74.76	68.08	65.43	sp|P21333	Filamin-A (	HUMAN	194
2	0	256.81	73.99	68.05	65.36	tr|Q5HY54	Filamin-A (	HUMAN	193
2	0	256.67	74.66	68.13	65.46	tr|Q60FE5	Filamin A (	HUMAN	193
3	252.92	252.92	71.17	60.78	56.1	sp|Q0966(	Neuroblast	HUMAN	185
4	231.83	245.31	76.63	72.94	70.48	sp|O7536	Filamin-B (	HUMAN	181
5	202.26	202.26	68.16	58.72	55.36	sp|P35579	Myosin-9 (	HUMAN	174
6	201.67	201.67	78.63	75.1	73.67	sp|Q0061(	Clathrin he	HUMAN	152
6	0	201.53	78.44	74.93	73.5	tr|A0A087	Clathrin he	HUMAN	152
7	190.94	190.94	67.68	59.87	57.2	sp|Q1381	Spectrin al	HUMAN	119
8	189.67	189.67	56.06	48.11	42.66	sp|P78527	DNA-depe	HUMAN	148
9	184.3	184.3	70.8	65.8	63.32	sp|Q9Y49(	Talin-1 OS=	HUMAN	120
10	176.97	176.97	66.55	57.59	54.8	sp|P49327	Fatty acid	HUMAN	112
11	171.48	171.48	79.83	70.3	70.3	sp|P08238	Heat shock	HUMAN	191
12	167.69	167.69	81.54	75.89	75.89	sp|P14618	Pyruvate k	HUMAN	176
13	166.69	166.69	90.15	88.66	87.46	sp|P04406	Glyceralde	HUMAN	208
14	166.04	166.04	69.08	59.18	52.28	sp|Q0108	Spectrin be	HUMAN	99
15	152.83	152.83	85.14	84.68	84.68	sp|P07437	Tubulin be	HUMAN	153
16	146.55	146.55	88.47	87.8	84.7	sp|P68363	Tubulin alp	HUMAN	137
17	146.26	146.26	94.4	89.07	81.6	sp|P60709	Actin, cyto	HUMAN	214
18	141.69	141.69	44.23	34.89	30.35	sp|Q1420	Cytoplasm	HUMAN	105
19	139.78	139.78	80.42	75.06	74.13	sp|P13639	Elongation	HUMAN	126
20	137.15	137.65	83.64	79.36	77.39	sp|O4370	Alpha-actir	HUMAN	107
21	136.72	136.72	89.27	82.19	78.11	sp|P08670	Vimentin C	HUMAN	147
22	134.35	134.35	71.76	67.46	63.69	sp|P07814	Bifunctiona	HUMAN	82
23	131.38	131.38	80.09	73.38	71.65	sp|P68104	Elongation	HUMAN	151
24	130.95	130.95	57.07	50.37	44.9	sp|O7564	U5 small n	HUMAN	86
25	130.83	130.83	56.16	47.14	41.71	sp|Q9261(	Translatior	HUMAN	83
26	129.26	129.26	83.87	78.11	78.11	sp|P06733	Alpha-enol	HUMAN	155
27	127.31	127.31	70.8	63.41	57.68	sp|P42704	Leucine-ric	HUMAN	79
28	127.12	127.12	81.85	78.42	78.42	sp|Q1497	Importin s	HUMAN	102
29	124.92	124.92	80.95	75.8	73.33	sp|P55060	Exportin-2	HUMAN	97
30	120.56	120.56	81.39	76.8	75.81	sp|P55072	Transitiona	HUMAN	107
31	120.52	120.52	72.12	70.04	66.82	sp|P22314	Ubiquitin-l	HUMAN	85
32	117.57	117.57	79.26	75.7	74.46	sp|P11142	Heat shock	HUMAN	142
33	115.19	115.19	41.29	32.14	28.97	sp|Q7Z6Z7	E3 ubiquiti	HUMAN	86
34	111.97	111.97	56.75	50.28	46.44	tr|F8VPD4	CAD protei	HUMAN	73
34	0	111.25	55.42	48.85	45.12	sp|P27708	CAD protei	HUMAN	73
35	111.4	111.4	66.32	56.91	52.2	sp|P46940	Ras GTPas	HUMAN	63
36	108.13	137.57	59.67	48.13	43.22	sp|P35580	Myosin-10	HUMAN	90
37	104.2	104.2	67.24	62.91	57.56	sp|Q0821	ATP-depen	HUMAN	76
38	104.01	104.01	71.9	69.56	60.32	sp|O1498(	Exportin-1	HUMAN	64
39	103.43	103.43	68.62	60.41	53.33	sp|Q86VP(	Cullin-asso	HUMAN	64
40	101.96	101.96	61.58	49.45	48.73	sp|Q0083	Heterogen	HUMAN	87
41	101.03	101.03	52.46	45.78	39.31	sp|Q6P2Q	Pre-mRNA	HUMAN	68
42	100.44	100.44	69.71	66.63	62.89	sp|Q1320(	26S protea	HUMAN	65

Supplementary table 9

Click here to download Supplementary Material (for online publication): Supp Table 9.xlsx

A375 null and WT - all detected peptides covering potential mutant position

GDVEVSGPK
VATYDKLEK
LLSDLLPPSTGTFQEAQSR
EGEDPQASAQDETPITSAK
TTGIVMDSGDGVSHTVPIYEGYALPHAILR
TTGIVMDSGDGVSHTVPIYEGY
NFGAENPDPFVPVLSTAVK
GISEDSHLESLQDVGQSAAPTFMISPETITGTGK
TVASPGVSVEEAVEQIDIGGVTLLR
KADLINR
ASLEAAIADAEQHGELAIK
LDSTDFTSTIK
KAEGAPNQGK
NSFGLAPAAPLQVHAPLSPNQTVEISLPLSTVGSVMK
EAAEGLGSHER
SALFAQINQGESITHALK
SIDPGLKEDTLEFLIK
FLGVQDIVVGEGTHFLIPWVQKPIIFDCR
LECGGSAVEDK
DVPGFLQQSQSSGPGQPAVWHR
AWLSSQAAELER
GVVDSEEIPLNLSR
CDPGALVIPF
LGRVTIAQGGLLPNIQAVLLPK
NVDCILLAR
NSLEFFTMLAQR
SADHPTLDK
DVISNTSDVIGTYEAADVAQK
ESYPHVKTVCDAAEK
LFAYSSWEVLR
LNATEEMLQQELLSR
ASLIEEDEPAEK
GFGTDEQAIIDCLGSCSNK
AFSEYLGTDQSK
YQEEPPAPQPK
FLGQILTAFPALR
AFSSLNTLPEELRPYVPLLCSVLTK
ALDRPYTSK
AYLEGTCVEWLR
ILGEVINALK
GFSDKLDFLEGDQKPLAQCK
VTVHKNK
VLHASLQSVLHK
EALDVLDAVLK
DTEGGPKEEESPV
AAAAAAAAAPAAAATAATTAATTAATAAQ
IEKEEQDKKR

Supplementary table 10

Click here to download Supplementary Material (for online publication): Supp Table 10.xlsx

Peptide Name	Sequence	no of AAs	2D LC fraction (% water)	Precursor ion charge state
1	ASLEAAIADAEQ**H**GELAIK	19	79, **83**	2+, **3+**
2	WEAAHVAEQLR	11	79, **83**	**3+**
3	YFAGNLASGGAAGATSLCFVYPLDFAR (new trypsin cleavage site before Y)	27	50, **65**	**3+**
4	QR**V**DEFEAL	9	**91**	**2+**
5	**V**DEFEAL	7	91, **95**	**2+**
6	SIITYVSSLYD**T**MPR	15	**65**	2+, **3+**
7	LII**N**LVTSIGDVVNHDPVVGDR	22	50**, 65**	2+, **3+**
8	GTAAAAAAAAAA**AA**K	15	75, **79**	**2+**, 3+
9	VSGSPEQAVEENLSSY**F**LDR	20	**79**	**3+**
10	ILGEVI**N**ALK	10	20, 50**, 65**	**2+**

Green = isotope labelled AA

Bold = mutation place

Highlighted = frac highlighted = m

Supplementary table 11

Click here to download Supplementary Material (for online publication): Supp Table 11.xls

DNAseq_hg38_VarScan2_p005_cov1

Chrom	Index	Chrom_id	Position	Ref	Var
NC_000001.11	1-69511-AG	1	69511	A	G
NC_000001.11	1-942451-TC	1	942451	T	C
NC_000001.11	1-953279-TC	1	953279	T	C
NC_000001.11	1-964350-TG	1	964350	T	G
NC_000001.11	1-973858-GC	1	973858	G	C
NC_000001.11	1-976215-AG	1	976215	A	G
NC_000001.11	1-978953-CG	1	978953	C	G
NC_000001.11	1-979472-GC	1	979472	G	C
NC_000001.11	1-979496-TC	1	979496	T	C
NC_000001.11	1-981169-AG	1	981169	A	G
NC_000001.11	1-999842-CA	1	999842	C	A
NC_000001.11	1-1044368-AT	1	1044368	A	T
NC_000001.11	1-1072052-GA	1	1072052	G	A
NC_000001.11	1-1334174-TC	1	1334174	T	C
NC_000001.11	1-1490373-TC	1	1490373	T	C
NC_000001.11	1-1534019-TC	1	1534019	T	C
NC_000001.11	1-1616547-TC	1	1616547	T	C
NC_000001.11	1-1640326-GA	1	1640326	G	A
NC_000001.11	1-1668373-CT	1	1668373	C	T
NC_000001.11	1-1719406-GA	1	1719406	G	A
NC_000001.11	1-1734736-CT	1	1734736	C	T
NC_000001.11	1-1734812-GA	1	1734812	G	A
NC_000001.11	1-1754601-GT	1	1754601	G	T
NC_000001.11	1-1923929-CT	1	1923929	C	T
NC_000001.11	1-1927692-CT	1	1927692	C	T
NC_000001.11	1-1956754-CA	1	1956754	C	A
NC_000001.11	1-1968747-TC	1	1968747	T	C
NC_000001.11	1-1968793-TC	1	1968793	T	C
NC_000001.11	1-2502461-CT	1	2502461	C	T
NC_000001.11	1-2509919-TC	1	2509919	T	C
NC_000001.11	1-2512975-GA	1	2512975	G	A
NC_000001.11	1-2556714-AG	1	2556714	A	G
NC_000001.11	1-2595307-AG	1	2595307	A	G
NC_000001.11	1-3411794-TC	1	3411794	T	C
NC_000001.11	1-3412703-GA	1	3412703	G	A
NC_000001.11	1-3473163-CT	1	3473163	C	T
NC_000001.11	1-3499885-CA	1	3499885	C	A
NC_000001.11	1-3505333-AG	1	3505333	A	G
NC_000001.11	1-3579915-TC	1	3579915	T	C
NC_000001.11	1-3761369-TC	1	3761369	T	C
NC_000001.11	1-3836572-AT	1	3836572	A	T
NC_000001.11	1-3891029-GC	1	3891029	G	C
NC_000001.11	1-6124032-AG	1	6124032	A	G
NC_000001.11	1-6245232-CA	1	6245232	C	A
NC_000001.11	1-6253878-CT	1	6253878	C	T
NC_000001.11	1-6418980-GA	1	6418980	G	A
NC_000001.11	1-6554331-AC	1	6554331	A	C
NC_000001.11	1-6554355-AG	1	6554355	A	G
NC_000001.11	1-6554475-GA	1	6554475	G	A
NC_000001.11	1-6575171-AG	1	6575171	A	G
NC_000001.11	1-6645884-CT	1	6645884	C	T
NC_000001.11	1-6887657-CT	1	6887657	C	T
NC_000001.11	1-7737443-CG	1	7737443	C	G
NC_000001.11	1-7820623-TG	1	7820623	T	G
NC_000001.11	1-7827433-TC	1	7827433	T	C

Supplementary table 12

Click here to download Supplementary Material (for online publication): Supp Table 12.xls

Index11_hg38_cov10_p005

Chrom	Position	Ref	Var	Index
1	953279	T	C	1-953279-TC
1	1044368	A	T	1-1044368-AT
1	1490373	T	C	1-1490373-TC
1	1534019	T	C	1-1534019-TC
1	1623412	T	C	1-1623412-TC
1	1668373	C	T	1-1668373-CT
1	1719406	G	A	1-1719406-GA
1	1734736	C	T	1-1734736-CT
1	1754601	G	T	1-1754601-GT
1	3891029	G	C	1-3891029-GC
1	6253878	C	T	1-6253878-CT
1	6554331	A	C	1-6554331-AC
1	6633037	A	G	1-6633037-AG
1	6645884	C	T	1-6645884-CT
1	6887657	C	T	1-6887657-CT
1	9744635	C	T	1-9744635-CT
1	11126694	G	A	1-11126694-GA
1	12022869	A	G	1-12022869-AG
1	15427166	G	A	1-15427166-GA
1	15506048	T	C	1-15506048-TC
1	15807483	T	C	1-15807483-TC
1	15807492	T	C	1-15807492-TC
1	15929512	T	C	1-15929512-TC
1	16251299	C	T	1-16251299-CT
1	16938425	C	T	1-16938425-CT
1	16939065	C	T	1-16939065-CT
1	16986248	C	T	1-16986248-CT
1	19238850	C	G	1-19238850-CG
1	20650507	A	C	1-20650507-AC
1	20661380	G	C	1-20661380-GC
1	20765422	C	T	1-20765422-CT
1	20941500	G	C	1-20941500-GC
1	21880156	T	C	1-21880156-TC
1	22509184	A	G	1-22509184-AG
1	22520216	G	A	1-22520216-GA
1	23092790	G	A	1-23092790-GA
1	23559007	T	C	1-23559007-TC
1	23854472	T	C	1-23854472-TC
1	26284400	A	C	1-26284400-AC
1	26557020	A	C	1-26557020-AC
1	26884230	T	C	1-26884230-TC
1	26993865	T	C	1-26993865-TC
1	27838846	A	C	1-27838846-AC
1	27882855	A	G	1-27882855-AG
1	29148829	C	T	1-29148829-CT
1	29148882	C	G	1-29148882-CG
1	29149136	T	G	1-29149136-TG
1	29216125	A	G	1-29216125-AG
1	32695611	T	C	1-32695611-TC
1	34981997	C	G	1-34981997-CG
1	35097364	G	A	1-35097364-GA
1	35760347	T	C	1-35760347-TC
1	35760519	C	A	1-35760519-CA
1	36092018	G	A	1-36092018-GA
1	36286832	C	T	1-36286832-CT

Supplementary table 13

Click here to download Supplementary Material (for online publication): Supp Table 13.xlsx

NO	unique mutant entry
1	PLEC_ENST00000322810_A2194V_MUT
2	PLEC_ENST00000322810_A2194V_MUT
3	AHNAK_ENST00000378024_D3312E_MUT
4	MYH9_ENST00000216181_I1626V_MUT
5	MYH9_ENST00000216181_I1626V_MUT
6	DYNC1H1_ENST00000360184_R462H_MUT
7	TLN1_ENST00000314888_S1227L_MUT
8	HSPA8_ENST00000534624_V602I_MUT
9	TUBA1C_ENST00000541364_T410S_MUT
10	EEF1A1_ENST00000331523_M335T_MUT
11	EEF1A1_ENST00000331523_M335T_MUT
12	NUMA1_ENST00000620566_G2046E_MUT
13	CCT8_ENST00000286788_S380A_MUT
14	ECPAS_ENST00000338205_N698S_MUT
15	HSPH1_ENST00000320027_L165I_MUT
16	HSPH1_ENST00000320027_L165I_MUT
17	AKAP12_ENST00000402676_E1600D_MUT
18	AKAP12_ENST00000402676_E1600D_MUT
19	XPO1_ENST00000406957_T960R_MUT
20	TGM2_ENST00000361475_G208V_MUT
21	MYBBP1A_ENST00000254718_Q8E_MUT
22	HEATR1_ENST00000366582_N1694S_MUT
23	EIF4G1_ENST00000441154_A554P_MUT
24	PABPC1_ENST00000318607_R481C_MUT
25	NES_ENST00000368223_V130A_MUT
26	MSN_ENST00000360270_P86H_MUT
27	GPI_ENST00000415930_I247T_MUT
28	KIF5B_ENST00000302418_V409D_MUT
29	LARS_ENST00000394434_N892D_MUT
30	ATIC_ENST00000236959_T116S_MUT
31	MAP4_ENST00000360240_S427Y_MUT
32	LAMC1_ENST00000258341_I458V_MUT
33	PTBP1_ENST00000635647_L435I_MUT
34	PTBP1_ENST00000635647_L435I_MUT
35	RRP12_ENST00000414986_R1220Q_MUT
36	EIF3B_ENST00000397011_S64P_MUT
37	EIF5B_ENST00000617677_K522T_MUT
38	RRBP1_ENST00000377813_G449D_MUT
39	PDIA3_ENST00000300289_L480P_MUT
40	DNM1L_ENST00000549701_S563I_MUT
41	CCT7_ENST00000258091_Q17K_MUT
42	IARS2_ENST00000366922_L794I_MUT
43	PDCD11_ENST00000369797_L1859P_MUT
44	PGM2_ENST00000381967_A2S_MUT
45	RNPEP_ENST00000295640_P198T_MUT
46	ENO3_ENST00000518175_S362T_MUT
47	ENO3_ENST00000518175_S362T_MUT
48	TRAP1_ENST00000538171_D342E_MUT
49	GEMIN5_ENST00000285873_L1310I_MUT

Supplementary table 14

Click here to download Supplementary Material (for online publication): Supp Table 14.xlsx

NO	unique mutant entry
1	PLEC_ENST00000322810_A398T
2	PLEC_ENST00000322810_A398T
3	MYH9_ENST00000216181_I1626V
4	MYH9_ENST00000216181_I1626V
5	TLN1_ENST00000314888_S1227L
6	DYNC1H1_ENST00000360184_R462H
7	TUBA1C_ENST00000541364_T410S
8	HSPA8_ENST00000534624_S613G
9	EEF1A1_ENST00000331523_M335T
10	EEF1A1_ENST00000331523_M335T
11	EEF1A1_ENST00000331523_M335T
12	TGM2_ENST00000361475_G208V
13	TKT_ENST00000462138_P307A
14	AKAP12_ENST00000402676_E1600D
15	AKAP12_ENST00000402676_E1600D
16	MYBBP1A_ENST00000254718_Q8E
17	PTBP1_ENST00000349038_L378I
18	PTBP1_ENST00000349038_L378I
19	GPI_ENST00000415930_I247T
20	NES_ENST00000368223_V130A
21	EIF4G1_ENST00000441154_A554P
22	MSN_ENST00000360270_P86H
23	ATIC_ENST00000236959_T116S
24	MAP4_ENST00000360240_S427Y
25	LONP1_ENST00000590729_Q547E
26	LARS_ENST00000394434_N892D
27	LAMC1_ENST00000258341_I458V
28	EIF5B_ENST00000617677_K522T
29	IMMT_ENST00000449247_R708L
30	HNRNPK_ENST00000376281_V392G
31	NCAPD2_ENST00000315579_Q83E
32	CCT7_ENST00000258091_Q17K
33	AKR1B1_ENST00000285930_I205V
34	EIF3B_ENST00000397011_S64P
35	USP7_ENST00000344836_E604Q
36	XPO5_ENST00000265351_P571L
37	CAPN2_ENST00000295006_D22E
38	CAPN2_ENST00000295006_D22E
39	RPA1_ENST00000254719_S149P
40	RRBP1_ENST00000377813_L1043H
41	CAP1_ENST00000372805_S256A
42	FTSJ3_ENST00000427159_Q91E
43	FTSJ3_ENST00000427159_Q91E
44	PRPF6_ENST00000266079_G323E
45	PRRC2C_ENST00000426496_A906T
46	ARHGEF2_ENST00000313695_G509W
47	PGM2_ENST00000381967_A2S
48	PYGB_ENST00000216962_K622N
49	PHB_ENST00000617874_R43L

***Credit Author Statement**

Jakub Faktor: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing, Visualization
Giuseppa Grasso: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing
Filip Zavadil: Methodology, Software, Formal analysis, Writing, Visualization
Małgorzata Kurkowiak: Methodology, Software
Marcos Yébenes Mayordomo: Methodology, Software, Formal analysis, Writing
Sachin Kote: Methodology, Validation, Investigation
Ashita Singh: Methodology, Validation, Investigation
Li Ruidong: Methodology, Validation, Investigation, Writing
J. Robert O'Neill: Conceptualization
Petr Muller: Methodology, Visualization
David Goodlett: Conceptualization, Writing
Borek Vojtesek: Conceptualization, Methodology, Writing
Ted Hupp: Conceptualization, Methodology, Validation, Investigation, Writing, Visualization

***Declaration of Interest Statement**

Click here to download Declaration of Interest Statement: declaration-of-competing-interests.docx